Computational analysis of molecular vibrational spectra

Manuel Ferreras Moreno

A thesis presented for the degree of

Doctor of Philosophy in Chemistry

School of Physical And Chemical Sciences & School of Forestry

University of Canterbury

September, 2020

A mis Padres.

Acknowledgments

Firstly, my sincerest gratitude to my supervisor, Deb Crittenden, for the opportunity to live this incredible adventure. Her patience, support, knowledge and diligence have far exceeded all my expectations. I also wish to extend my warmest thank to Clemens Altaner, without whom this thesis would have been impossible.

Next, special thanks go to my wonderful kiwi family, for all their love and support: to Michael, for being the brother I never had but always dreamt of; to Nick, Vanessa and Tyler, for opening the doors of their home and making me feel part of it; to Ash, for leaving me now and taking away the biggest part of me; to Euan, for taking a little time to think things over; to Kevin, for those endless conversations; to Chris and Arman, for listening to every last question and being endlessly helpful; to Grant, for being a crazy horse; to Niki, for being an awesome companion and flatmate; to all the rest, for all the good and bad times from which I have learnt so much.

And, last but not least, to Sophia, for being the light when everything around me was darkness. Love.

From the bottom of my heart, THANKS.

* * *

En primer lugar, quiero agradecer a mi padre, madre y hermana todo el amor y el apoyo que siempre me han dado. Siempre. No creo que haya nadie mas afortunado que yo. En especial quiero agradecer a mis sobrinos, Lola y Álvaro, por todas esas llamadas que tanta fuerza me

dieron y tan querido me hicieron sentir aun estando a mas de 20000 km de distancia. Nunca podréis imaginar cuanto me ayudaron. Gracias.

También quiero agradecer a mis mejores amigos todo el apoyo: a mi Vecino, por hacerme sentir que nunca me fui, que siempre estuve a su lado; a Paco, por ser siempre mi muro de las lamentaciones; a Manolo, por ser mi compañero de viaje al centro del alma; y a Luis, por ser parte de mi aventura y aparecer cuando más lo necesitaba. Os quiero.

Desde lo más profundo de mi corazón, GRACIAS.

Contents

Chapter 1

Introduction

"...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Richard P. Feynman

Spectroscopic measurements are a foundational tool in chemical research that quantify the interaction between electromagnetic radiation and quantum states of matter. The details of experimental spectra are fundamentally determined by the electronic structure of the system under study, which is the ultimate basis of all its chemical properties. The utility of a spectrum lies in being able to extract detailed information about the way in which nuclei and electrons are organized within the system.

Molecular spectroscopy comprises a set of techniques for obtaining information on the electronic and geometric structure of molecules, both isolated and in different environments, providing information of the systems studied under the measurement conditions in which they have been recorded. Among the most widely used spectroscopic techniques we have *ultraviolet-visible (UV-vis) spectroscopy*, which interrogates transitions between electronic states of molecules; *microwave spectroscopy*, which measures the energy required to transition between different rotational quantum states; *nuclear magnetic resonance spectroscopy*, a technique used mainly in the elucidation of molecular structures, that is based on the absorption of electromagnetic radiation (radio frequency waves) by some atomic nuclei under the influence of a magnetic field; and *vibrational spectroscopy*, the focus of this study, which directly provides information on transitions between vibrational energy levels, and indirectly reflects the composition, structure and bonding of a system. The most common vibrational spectroscopic techniques are Infrared (IR) spectroscopy and Raman spectroscopy. An overview of the principles behind both of these techniques is provided in Appendix A.

The vibrational spectrum is a unique and characteristic physical property of a molecule or material. Thus, vibrational spectra can be used as a "molecular fingerprint" in the identification of unknown samples by comparison with reference spectra. Some examples of this can be found in the characterization and identification of polymers[1-3] as well as their structural and

surface properties[4,5]. Also, in the field of biological and medical sciences, the uses of vibrational spectroscopy are widespread in the characterization of lipids, nucleic acids, proteins and peptides, in addition to characterizing disease in animal tissues, recognizing plants and differentiating microbial cells, among others.[6-11] However, in recent times, new industrial and environmental applications of vibrational spectroscopy have been introduced. Industrial applications include the characterization of pharmaceutical materials for production processes,[12-14] food quality analysis,[14,15] identification of the different components of paints[16], pulp and paper quality control[17] and even potentially quantifying strain in wood.[18,19] Environmental applications include air, water and soil analysis, as well as measurement of gases and atmospheric compositions, which is crucial for understanding global climate changes and astronomical observations.[20-25]

However, if reference spectra are unavailable, interpretation of vibrational spectra is much harder. In some cases, vibrations arising from specific functional groups can be assigned using heuristic rules,[26] but in general, computational support is required to ensure vibrational spectra can be thoroughly and meaningfully assigned.[27]

Quantum chemical software packages provide access to a range of electronic structure models that can be used to describe how the energy changes as a function of molecular configuration (i.e. construct potential energy surfaces) and consequently solve the nuclear vibrational Schrödinger equation to predict vibrational spectra. A hierarchy of electronic structure models of different computational cost and accuracy are available, and these may be combined with a range of different models for predicting nuclear vibrational motion, as well, to yield overall predictions of varying accuracy and computational cost. A brief description of all electronic structure and nuclear vibrational models employed in this thesis is provided in the Methods section.

For small molecules, very accurate predictions of vibrational spectra are possible by constructing potential energy surfaces at high levels of *ab initio* theory and using nuclear vibrational models that account for anharmonicity;[28] the propensity of a molecule to spend more time at longer bond lengths to avoid the repulsive forces that are magnified upon bond compression, and also the way that stretching/compressing a molecule in one way makes it harder/easy to stretch/compress it in a different direction.

Because benchmark results can be generated for small molecules, they provide an ideal testing ground for trialling approximate and less computationally intensive schemes for constructing potential energy surfaces and/or solving the nuclear vibrational Schrodinger equation. In the first results chapter of this thesis (Chapter 3), we address the problem of finding more efficient schemes for constructing potential energy surfaces by taking a "two-level" approach – computing the harmonic part of the potential energy surface that describes symmetric vibrations along normal mode coordinates at a high level of *ab initio* theory, while constructing the remainder of the potential energy surface (the anharmonic region) at a lower level of theory.

The remainder of this thesis focusses on using more approximate nuclear vibrational models to interpret experimental observations. First, we study the nature of two novel chloride hydrates trapped in cyclopropenium chloride crystal. These chloride-water clusters are of fundamental importance in understanding a broad range of physicochemical processes in nature, particularly in biological and geological systems.[29-31] The discreteness of these clusters can be assessed by comparing experimental IR spectra to predicted IR spectra obtained from gas phase calculations. If a cluster is discrete, there will not be any specific interactions stabilizing it, so a gas phase model should provide quite accurate predictions. However, in some cases, specific interactions with the surrounding environment may contribute to forming and/or stabilising a particular chloride hydrate structural motif. In such cases, modelling the IR spectrum is more

complicated because environmental effects need to be taken into account. In Chapters 4 & 5, we investigate the structure and vibrational spectroscopy of two quite different chloride hydrate clusters that form within subtly different cyclopropenium chloride crystal environments.

For very large systems, applying quantum nuclear vibrational models can become impractical. In these cases, it may be appropriate to use classical physics-based approaches to modelling macromolecular structure and dynamics. In *Molecular Mechanics (MM)*, the energy of the system is then calculated as a function of the nuclear coordinates with the use of *force fields,* in which atoms are simulated as balls, and bonds as springs. To simulate the time-dependent evolution of the system, we can use Molecular Dynamics (MD) simulations. Further details of how force fields are defined and parameterised is provided in the Methods section, along with an explanation of how molecular dynamics simulations work and how key thermodynamic parameters such as temperature, pressure and volume are set and controlled.

Cellulose, the most abundant material in the biosphere,[32] is one such system that is not particularly amenable to quantum nuclear vibrational analysis.[33] The interest in this polymer has grown in the recent years as it has become a potential source of renewable fuel and materials.[34-38] The presence of hydrogen bonds play an important role, not only in the physical properties, solubility, hydroxyl reactivity and crystallinity, but also in the mechanical properties of cellulose.[39,40] The structure and hydrogen bonding patterns within common allotropes of cellulose (Iα and Iβ) have been studied experimentally and confirmed computationally.[41-47] Although some quantum harmonic frequency calculations have been performed,[45] they are not accurate enough to allow experimental IR spectra to be fully understood or assigned. Thermal shifts in IR stretching frequencies have also been observed.[48,49] Paradoxically, these may be easier to understand and explain than the IR frequencies themselves, because thermal effects can be modelled via classical molecular

dynamics simulations.[50,51] In the last results chapter (Chapter 6), we address this problem, using a classical molecular dynamics model to explore how changes in temperature affect O-H stretching vibrations and hydrogen bond lengths, and how this is coupled to and driven by changes along other vibrational modes of crystalline cellulose.

References

(1) Bower, D. I.; Maddams, W.: *The vibrational spectroscopy of polymers*; Cambridge University Press, 1992.

(2) Chalmers, J. M. Infrared spectroscopy in analysis of polymers and rubbers. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation* **2006**.

(3) Chalmers, J. M.; Hannah, R. W.; Mayo, D. W. Spectra–structure correlations: polymer spectra. *Handbook of Vibrational Spectroscopy* **2006**.

(4) Garton, A.: *Infrared spectroscopy of polymer blends, composites and surfaces*; Hanser New York, **1992**.

(5) Moskala, E.; Varnell, D.; Coleman, M. Concerning the miscibility of poly (vinyl phenol) blends—FTi. r. study. *Polymer* **1985**, *26*, 228-234.

(6) Lawson, E.; Barry, B.; Williams, A.; Edwards, H. Biomedical applications of Raman spectroscopy. *Journal of Raman Spectroscopy* **1997**, *28*, 111-117.

(7) Krafft, C.; Sergo, V. Biomedical applications of Raman and infrared spectroscopy to diagnose tissues. *Spectroscopy* **2006**, *20*, 195-218.

(8) Gremlich, H.-U.; Yan, B.: *Infrared and Raman Spectroscopy of Biological Materials*; CRC press, **2000**.

(9) Kalasinsky, V. F. Biomedical applications of infrared and Raman microscopy. *Applied Spectroscopy Reviews* **1996**, *31*, 193-249.

(10) Petrich, W. Mid-infrared and Raman spectroscopy for medical diagnostics. *Applied Spectroscopy Reviews* **2001**, *36*, 181-237.

(11) Mantsch, H. H.; Chapman, D.: *Infrared spectroscopy of biomolecules*; Wiley-Liss New York, 1996.

(12) Clark, D.; Pysik, A. The analysis of pharmaceutical substances and formulated products by vibrational spectroscopy. *Handbook of Vibrational Spectroscopy* **2006**.

(13) Bugay, D. E. Characterization of the solid-state: spectroscopic techniques. *Advanced Drug Delivery Reviews* **2001**, *48*, 43-65.

(14) Broad, N.; Graham, P.; Hailey, P.; Hardy, A.; Holland, S.; Hughes, S.; Lee, D.; Prebble, K.; Salton, N.; Warren, P. Guidelines for the development and validation of near-infrared spectroscopic methods in the pharmaceutical industry. *Handbook of Vibrational Spectroscopy* **2002**, *5*, 3590-3610.

(15) Reddy Gangidi, R.; Proctor, A.: Chapter 8 - Meat and Meat Products. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, D.-W., Ed.; Academic Press: San Diego, **2009**; pp 179-214.

(16) Urban, M. W. Infrared and R aman Spectroscopy and Imaging in Coatings Analysis. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation* **2006**.

(17) Griffiths, P. R.; Chalmers, J. M.: *Handbook of Vibrational Spectroscopy*; Wiley Online Library, **2002**.

(18) Guo, F.; Altaner, C. M. Measuring Molecular Strain in Rewetted and Never-Dried Eucalypt Wood with Raman Spectroscopy. *Biomacromolecules* **2019**, *20*, 3191-3199.

(19) Guo, F.; Cramer, M.; Altaner, C. M. Evaluation of near infrared spectroscopy to non-destructively measure growth strain in trees. *Cellulose* **2019**, *26*, 7663-7673.

(20) Griffith, D. W. FT-IR Measurements of Atmospheric Trace Gases and their Fluxes. *Handbook of vibrational spectroscopy* **2006**.

(21) Griffith, D. W.; Jamie, I. M. F ourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation* **2006**.

(22) Hofmeister, A. M.; Wopenka, B.; Locock, A. J. Spectroscopy and structure of hibonite, grossite, and CaAl2O4: Implications for astronomical environments. *Geochimica et Cosmochimica Acta* **2004**, *68*, 4485-4503.

(23) Klemperer, W. Astronomical chemistry. *Annual review of physical chemistry* **2011**, *62*, 173-184.

(24) Bacsik, Z.; Mink, J.; Keresztury, G. FTIR spectroscopy of the atmosphere. I. Principles and methods. *Applied Spectroscopy Reviews* **2004**, *39*, 295-363.

(25) Visser, T. Infrared spectroscopy in environmental analysis. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation* **2006**.

(26) Andreev, G. N.; Argirov, O. K.; Penchev, P. N. Expert system for the interpretation of infrared spectra. *Analytica Chimica Acta* **1993**, *284*, 131-136.

(27) Palafox, M. A. Computational chemistry applied to vibrational spectroscopy: A tool for characterization of nucleic acid bases and some of their 5-substituted derivatives. *Physical Sciences Reviews* **2017**, *2*, 20160132.

(28) Sibaev, M.; Crittenden, D. L. PyVCI: A flexible open-source code for calculating accurate molecular infrared spectra. *Computer Physics Communications* **2016**, *203*, 290-297.

(29) Castleman, A.; Bowen, K. Clusters: Structure, energetics, and dynamics of intermediate states of matter. *The Journal of Physical Chemistry* **1996**, *100*, 12911-12944.

(30) Ohtaki, H.; Radnai, T. Structure and dynamics of hydrated ions. *Chemical Reviews* **1993**, *93*, 1157-1204.

(31) Roux, B.; Bernèche, S.; Egwolf, B.; Lev, B.; Noskov, S. Y.; Rowley, C. N.; Yu, H. Ion selectivity in channels and transporters. *Journal of General Physiology* **2011**, *137*, 415-426.

(32) Dixon, R. K.; Brown, S.; Houghton, R. A.; Solomon, A. M.; Trexler, M. C.; Wisniewski, J. Carbon Pools and Flux of Global Forest Ecosystems. *Science* **1994**, *263*, 185-190.

(33) Agarwal, V.; Huber, G. W.; Jr., W. C. C.; Auerbach, S. M. Simulating infrared spectra and hydrogen bonding in cellulose Iβ at elevated temperatures. *The Journal of Chemical Physics* **2011**, *135*, 134506.

(34) Moon, R. J.; Martini, A.; Nairn, J.; Simonsen, J.; Youngblood, J. Cellulose nanomaterials review: structure, properties and nanocomposites. *Chemical Society Reviews* **2011**, *40*, 3941-3994.

(35) Pattinson, S. W.; Hart, A. J. Additive Manufacturing of Cellulosic Materials with Robust Mechanics and Antimicrobial Functionality. *Advanced Materials Technologies* **2017**, *2*, 1600084-n/a.

(36) W¸stenberg, T. Cellulose and cellulose derivatives in the food industry: fundamentals and applications. **2015**.

(37) Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* **2007**, *315*, 804-807.

(38) Carroll, A.; Somerville, C. Cellulosic Biofuels. *Annual Review of Plant Biology* **2009**, *60*, 165-182.

(39) Fan, M.; Dai, D.; Huang, B.: Fourier transform infrared spectroscopy for natural fibres. In *Fourier transform-materials analysis*; InTech, 2012.

(40) Altaner, C. M.; Thomas, L. H.; Fernandes, A. N.; Jarvis, M. C. How Cellulose Stretches: Synergism between Covalent and Hydrogen Bonding. *Biomacromolecules* **2014**, *15*, 791-798.

(41) Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. Neutron Crystallography, Molecular Dynamics, and Quantum Mechanics Studies of the Nature of Hydrogen Bonding in Cellulose Iβ. *Biomacromolecules* **2008**, *9*, 3133-3140.

(42) Nishiyama, Y.; Langan, P.; Wada, M.; Forsyth, V. T. Looking at hydrogen bonds in cellulose. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 1172-1177.

(43) Makarem, M.; Lee, C. M.; Kafle, K.; Huang, S.; Chae, I.; Yang, H.; Kubicki, J. D.; Kim, S. H. Probing cellulose structures with vibrational spectroscopy. *Cellulose* **2019**, *26*, 35-79.

(44) Maréchal, Y.; Chanzy, H. The hydrogen bond network in Iβ cellulose as observed by infrared spectrometry. *Journal of Molecular Structure* **2000**, *523*, 183-196.

(45) Lee, C. M.; Kubicki, J. D.; Fan, B.; Zhong, L.; Jarvis, M. C.; Kim, S. H. Hydrogen-Bonding Network and OH Stretch Vibration of Cellulose: Comparison of Computational Modeling with Polarized IR and SFG Spectra. *The Journal of Physical Chemistry B* **2015**, *119*, 15138-15149.

(46) Kubicki, J. D.; Mohamed, M. N.-A.; Watts, H. D. Quantum mechanical modeling of the structures, energetics and spectral properties of Iα and Iβ cellulose. *Cellulose* **2013**, *20*, 9-23.

(47) Li, Y.; Lin, M.; Davenport, J. W. Ab Initio Studies of Cellulose I: Crystal Structure, Intermolecular Forces, and Interactions with Water. *The Journal of Physical Chemistry C* **2011**, *115*, 11533-11539.

(48) Altaner, C. M.; Horikawa, Y.; Sugiyama, J.; Jarvis, M. C. Cellulose Iβ investigated by IR-spectroscopy at low temperatures. *Cellulose* **2014**, *21*, 3171-3179.

(49) Ivanova, N.; Korolik, E.; Zhbankov, R.; Insarova, N. Low-temperature IR spectra of cellulose. *Journal of Applied Spectroscopy* **1983**, *38*, 676-678.

(50) Bergenstråhle, M.; Berglund, L. A.; Mazeau, K. Thermal response in crystalline Iβ cellulose: a molecular dynamics study. *The Journal of Physical Chemistry B* **2007**, *111*, 9138-9145.

(51) Zhang, Q.; Bulone, V.; Ågren, H.; Tu, Y. A molecular dynamics study of the thermal response of crystalline cellulose Iβ. *Cellulose* **2011**, *18*, 207-221.

Chapter 2

Methods

"Chemistry is no longer a purely experimental science."

The Royal Swedish Academy of Science, awarding W.Kohn
and J.Pople the 1998 Nobel Prize in Chemistry

2.1 Introduction

Computational chemistry is a branch of chemistry that uses computers to study and solve chemical problems, by applying computational models to simulate molecular systems. It uses chemistry theories, concepts and mathematical models based on classical and quantum physics and statistical mechanics to calculating the structure and properties of molecules, molecular aggregates and materials.[1] There is no ideal model for all possible applications, and typically trade-offs have to be made between accuracy, ease of use, and computational cost. Nonetheless, computational models have been widely used to solve a wide variety of problems including predicting the binding of drugs to receptors[2-4], predicting phase diagrams[5-8], modelling the structures of molecules and nanoparticles[9-12], and simulating IR spectra[13-16].

From a classical point of view, the evolution of a system over time is determined by the interaction between particles and the conservation of energy required by the First Law of Thermodynamics. Interactions between particles are described in the field of *Molecular Mechanics* (MM) with the use of force fields, which are parametrized mathematical approximations to the potential energy surface (PES) of a system. These force fields describe how the energy of a system changes as a function of its molecular configuration, V(**R**), where **R** represents the 3 x N atom coordinate vector of atomic positions. In the same classical context, Molecular Dynamics (MD) describes how atomic positions change as a function of time **R**(t), using a pre-defined force field and solving Newton's equations of motion.

In the quantum approach, we aim to solve the many-particle Schrödinger equation. For most chemical systems, quantum dynamical effects do not occur on experimentally-accessible time scales, so the time-independent Schrödinger equation is the most appropriate form. The Born-Oppenheimer approximation[17] allows us to decouple electronic motions from nuclear motions.

Solutions to the electronic Schrödinger equation for different nuclear coordinates give a series of energies that allow us to construct the potential energy surface V(**R**). The methods used to solve the electronic Schrödinger equation are referred to *electronic structure methods*. Once we have the PES, we can use it to solve the nuclear vibrational Schrödinger equation and obtain the nuclear vibrational wavefunction Ψ(**R**), which describes the coupled vibrational motions of the atoms in the system, using different *quantum nuclear vibrational models*.

So, while MD and quantum nuclear vibrational models do both describe how nuclei move, they do it in substantially different ways: Ψ(**R**) is effectively a "standing wave" solution of the quantum nuclear vibrational Schrodinger equation, whereas **R**(t) also captures wave-like motions but in a time-dependent way. However, both require the potential energy surface V(**R**) as input, because both methods balance and interconvert kinetic and potential energy, which requires knowledge of the potential energy in the first place. The relationship between classical and quantum approaches to constructing potential energy surfaces, V(**R**), and modelling nuclear dynamics is summarised in Table 1.

	Classical	**Quantum**
V(R)	Molecular Mechanics Force Fields	Electronic Structure Calculations
R(t) / Ψ(**R**)	Molecular Dynamics	Quantum Nuclear Vibrational Models

Table 1. Quantum and classical approaches to expressing the potential energy surface (PES) and modelling nuclear vibrations.

It is important to clarify that each column of Table 1 represents the most common combination of methods: a classical description of the PES (force field) is most commonly used as the basis for classical molecular dynamics simulations, and a quantum description of the PES employed

when modelling quantum nuclear vibrations. However, other mixed approaches are also possible. For example, *on-the-fly MD*[18] evaluates energies quantum mechanically during molecular dynamics simulations. Similarly, many MD program packages[19-21] include code for *normal mode analysis*, which performs a quantum harmonic frequency calculation using a classical force field to obtain the required potential energy information.

The aim of this chapter is to provide an explanation of the classical and quantum computational methods presented above that were used during the materialization of this thesis. I have tried to describe them as simply as possible, keeping the use of mathematical formulas to a minimum. This will be followed by a brief rationalization of methodological choices made in this thesis as a whole. The next two sections will focus on methods for computing or expressing the potential energy surface, followed by methods for simulating classical motion and quantum nuclear vibrations. In each case, the quantum approach will be presented before its classical counterpart.

2.2. Electronic Structure Methods

In the context of quantum mechanics, the state of a system can be described as a function of the coordinates of its particles and time, Ψ, called the wavefunction. The wavefunction contains all the information about the system that can be known. Within a non-relativistic approximation, Ψ is obtained from the time-dependent Schrödinger equation. A subset of the solutions to the time-dependent Schrödinger equation correspond to standing waves. These are the most relevant for systems whose properties are determined by the distribution of quantum particles (e.g. electron densities) rather than the dynamical motions of

those particles (e.g. proton tunneling). They can be found by solving the simpler non-relativistic and time-independent Schrödinger equation:

$$\widehat{H}\Psi_n = E_n\Psi_n \tag{2.1}$$

In this differential equation, $\widehat{H}$ is the Hamiltonian operator associated with the energy of the electrons and nuclei of the system, Ψ_n is the wavefunction that characterizes the distribution of quantum particles of the system in its n^{th} quantum state, and E_n is the total energy of that state. The molecular Hamiltonian contains terms for the kinetic energy of nuclei and electrons, repulsion between nuclei, attraction between nuclei and electrons, and repulsion between electrons.

The Schrödinger equation can only be solved analytically for up to two-particle systems; therefore, the use of approximations is needed for larger systems. The *Born-Oppenheimer approximation* is based on the fact that the mass of the nuclei is thousands of times larger than the mass of the electrons and therefore, they move much more slowly. This simplifies the problem by splitting the Hamiltonian into two terms: nuclear and electronic. By doing this, the electronic motion can be described by taking into account the field created by the fixed nuclei. That allows us the solve the electronic and the nuclear Schrödinger equations separately.

The electronic Schrödinger equation is still analytically unsolvable for systems with more than one electron due to the interelectronic repulsion term. However, there exist different methods that attempt to solve it, called *electronic structure methods*[22,23], that are characterized by different levels of mathematical approximations. In this section, the electronic structure methods used in this thesis will be described. There are two categories: *wave function methods,* also referred to as *ab initio* methods, which are based on approximations made to the

wavefunction; and *density functional* methods, which compute the energy as a functional of the electronic density of the system.

At last, we will focus on *basis sets*, another key factor in the resolution of the electronic structure problem in both wavefunction and density functional methods. Basis sets are sets of mathematical functions that are used to construct the wavefunction (or the electron density in density functional methods). The choice of the basis sets should be strategic, in accordance with the system under study, the method used, the desired level of accuracy, and the computational time available.

2.2.1. Wavefunction methods – Hartree-Fock Theory

As has been mentioned above, the Schrödinger equation cannot be solved analytically for systems with more than one electron due to the interelectronic repulsion term. The first main approximation was introduced with the Hartree-Fock theory[22-25], in which the electronic wavefunction is approximated to an antisymmetrized product of one-electron functions, i.e. the motion of each electron is considered independent; although it experiences a time-averaged electrostatic potential created by all the other electrons. The Hartree-Fock wavefunction is then expressed as a Slater determinant[1], whose elements are molecular orbitals expressed as linear combinations of atomic orbital functions, which satisfy the Pauli Exclusion Principle. The combination of atomic orbitals that give the best wave function is found using the variational method, where their coefficients are optimized iteratively until the energy of the system converges to its minimum. This process is known as the self-consistent field (SCF) method or Hartree-Fock (HF) method.

The time-averaged electrostatic potential used in the HF method does not take into account the instantaneous repulsion interaction between electrons, it only satisfies the antisymmetry condition. Therefore, the HF wavefunction does not correlate the motions of opposite spin electrons. This *correlation energy*[26] is thus defined as the difference between the exact nonrelativistic energy and the Hartree-Fock energy. Various methods have been developed to calculate it. They will be described in the following section.

2.2.2. Wavefunction methods – Post Hartree-Fock

Post Hartree-Fock methods take the Hartree-Fock method as a starting point and introduce corrections to it by constructing new wave functions in order to account for the correlation energy.

Generally, an infinite number of Slater determinants are needed to construct an exact wavefunction for a system. The *configuration interaction (CI)* method[27,28] is based on the linear combination of several Slater determinants and involves optimizing their respective variable coefficients. These Slater determinants are constructed from the orbitals obtained during a HF calculation, some of which are optimized (occupied orbitals) and others of which are not (virtual orbitals). Invoking the variational principle which states that any approximate energy is an upper bound to the true energy when it is calculated as the expected value of a trial wave function and the true Hamiltonian, the coefficients of those newly included excitations are optimized. The CI wavefunctions are classified according to the number of electron excitations to the virtual orbitals allowed: single excited, double excited, etc. Therefore, a full CI calculation (FCI) includes all the possible configurations in a system and would provide exact solutions to the electronic Schrödinger equation for all quantum states of the system if the basis set used was complete. Unfortunately, FCI calculations are very rare due to the high

computational cost they require. Nonetheless, truncated CI models are frequently used to model both ground and excited states of molecular systems.

In the *Coupled Cluster (CC)* method[28]**,** the molecular wavefunction is expressed as the product of the HF ground state wavefunction and an *exponential excitation operator* or *cluster operator*. This operator is the sum of different operators, each of them accounting for the number of excited electrons. The purpose of this operator is to express the CC wavefunction as a combination of Slater determinants that include the HF ground state and all the possible electron excitations from occupied orbitals to virtual ones.

In practice, the cluster operator is truncated to finite order. Because chemistry is fundamentally about the behavior of pairs of electrons, the minimum practical truncation order is two. Including single and double excitations gives the *Coupled Cluster Singles and Doubles (CCSD) method.* Unlike in CI theory, the coefficients for each contributing determinant are not obtained variationally, but instead by solving a series of coupled equations. Including triple excitations (CCSDT) improves accuracy but at high additional computational cost[28]. An affordable alternative is *CCSD(T)*, which includes approximations for the triple excitations that make it applicable to molecules up to a moderate size[28].

An alternative approach to solving the electronic correlation problem is *Rayleigh-Schrödinger perturbation theory (RSPT)*[29]. RSPT was first applied in the electronic structure context by Møller and Plesset[30]. Using the standard RSPT approach of dividing the Hamiltonian into a solvable part (the Hartree-Fock Hamiltonian) and a perturbation term (the Coulomb operator), they then expanded the wavefunction adding excitations order by order, and also expanded the energy in a Taylor series. The perturbation expansion is most frequently truncated at 2nd order, yielding the MP2 energy correction expression.

Although this derivation procedure is formally quite different from CI and CC theories, it reduces to a series of energy contributions from excited-state determinants and their electron-repulsion weighted overlap with the ground state, divided by the difference in energy between the ground and excited state. Therefore, the more similar an excited state is to the ground state, both in terms of Coulomb-weighted electron distribution and energy, the larger the contribution it will make to the correlation energy.

It is important to note that, unlike CI and CC methods, this perturbational approach is not strictly variational,[31] because convergence of the perturbation theory expansion is not guaranteed. However, in practice, MP2 is variational because the perturbation theory is truncated before it diverges.[32]

An important consideration before finishing this section is to highlight the fact that there is a hierarchy of computational cost and accuracy among all the methods to calculate the electronic correlation energy previously described. Usually, variational approaches are more accurate and computationally expensive than perturbative ones. For instance, MP2 is a less accurate method than CISD and CCSD, despite all models including terms with up to double excitations of electrons from bonding orbitals to "virtual" orbitals. Therefore, choosing the method in accordance with the system under study is crucial.

2.2.3. Density Functional Theory

Density Functional Theory (DFT) is based on the *Hohenberg-Kohn theorem*[33], which states that *'the ground state energy of any interacting many particle system with a given fixed inter-particle interaction is a unique functional of the electron density* $\rho(r)$*'*. In other words,

the ground state energy of a system can be calculated from its electron density instead of the wave function[34]. The energy, $E(\rho)$, is a density functional, and can be expressed as follows:

$$E = E_T(\rho) + E_V(\rho) + E_R(\rho) + E_{XC}(\rho) \quad (2.2)$$

where E_T is again the non-interacting electron kinetic energy functional; E_V, the nucleus-electron attraction energy functional; E_R, the "classical" electron-electron repulsion energy functional (does not and cannot account for electrons exchanging between orbitals which don't formally exist within DFT); and E_{XC}, the exchange-correlation energy functional, which captures the missing terms – the exchange energy that arises from electrons of the same spin changing configurations, and the missing correlation from the Hartree-Fock model due to the mean-field approximation that applies to both same-spin and opposite-spin electron-electron interactions.

A functional is a function of a function that yields a single value, in this case, a procedure by which the energy is computed from the electron density. A practical application of this theory was developed by Kohn and Sham[34]. In their formulation, the electron density is obtained from a HF-like wavefunction by integrating over all coordinates of all electrons but one. However, the underlying orbitals are retained and used in refining the electron density, optimizing the energy by changing MO coefficients while applying a density functional model to predict the missing exchange-correlation energy.

The key problem with all density functional theories arises with the exchange-correlation energy since there are no universal functional forms to describe them. The accuracy of the results from DFT calculations will depend upon the ability of the functional to predict this exchange-correlation energy. There is an increasingly large number of functionals that have

been developed to address this problem, making use of different approximations and techniques[1,35-37].

In the *Local Density Approximation (LDA)*[38]*,* the exchange-correlation energy depends only on the electron density and each contribution is treated separately. The exchange energy is calculated taking an electron gas with homogeneous and constant electron density as a model. On the other hand, several approaches exist for the correlation energy term. The most sophisticated one is an analytical expression developed by Vosko, Wilk and Nusair[39], based on Monte Carlo calculations. In LDA approximations, exchange-correlation effects are assumed to be local and only dependent on the value of the electron density on each point. Generally, LDA methods provide surprisingly good results, especially if one considers the simplicity of the model on which they are based.

For open-shell molecules and molecular geometries close to disassociation, the *Local Spin-Density Approximation (LSDA)* gives better results than LDA[39]. In the LSDA approximation, contrary to LDA, electrons with opposite spin, paired with each other, are allowed to have different spatial orbitals. Therefore, electron densities due to positive spin electrons are treated separately from those with negative ones and functionals become dependent on both densities.

The next step is to introduce density gradients in the description of the exchange-correlations effects. The *Generalized-Gradient Approximation (GGA)* considers the value of the density at each point and how this density varies around each of them. Functionals that include this approximation are called semi-local because the energy per unit of volume continues to depend solely on the density and the derivative of the density at each point. One of the most common GGA approximations was developed by Becke[40], who added a correction term to the LDA exchange functional due to its usual 10-15% underestimation of this energy. For the correlation

energy functionals, the most usual corrections were developed by Perdew and Wang[41] (PW91); and Lee, Yang and Parr[42] (LYP).

Another way to tackle the problem is by using *hybrid functionals*. These functionals allow the calculation of the exchange-correlation energy by a combination of local and/or semi-local correlation functionals and an HF exact exchange energy. One of the most widely hybrid functionals is B3LYP[43], which uses local functionals that include spin polarization (LDSA) and semi-local functionals (Becke functional for the exchange and LYP for electronic correlation) together with a 20% of exact HF exchange energy contribution.

2.2.4. Basis Sets

A *basis set* is a set of functions (called *basis functions*) that is used to construct molecular orbitals and electron densities in wave function and density functional methods respectively. Basis sets are sets of functions that can be used to mathematically approximate atomic orbitals, and how they change in size in response to their molecular environment. Linear combinations of atomic orbitals are foundational to the Hartree-Fock and Kohn-Sham equations, transforming them from an unsolvable set of coupled differential equations into an algebraically solvable form[1]. The most commonly used atomic orbitals are *Slater-type orbitals (STOs)* and *Gaussian-type orbitals (GTOs)*[44].

Slater-type orbitals are solutions to the Schrödinger equation of hydrogen-like atoms. They accurately describe electron density near the nucleus and decay exponentially far away from it. However, two factors make STOs not suitable for most calculations: they are not able to describe electron correlation effects due to their hydrogen-like atomic nature, and they are

difficult to manage since they do not allow a fast resolution of the integrals involved in computational chemistry methods[45].

The solution to the above problem is to express Slater-type orbitals as a linear combination of Gaussian-type orbitals. Although GTOs represent the atomic orbitals less precisely than STOs, GTOs simplify the analytical resolution of integrals due to their mathematical expression, leading to huge computational savings. There is a myriad of Gaussian-type orbital basis sets in the literature, which come in hierarchies of increasing size and properties[44].

A *minimal basis set*[1] contains the minimum number of basis functions that are needed for each atom in the molecule, that is, a basis function for each atomic orbital in the ground state of the corresponding free atom. Common minimal basis sets are expressed as STO-nG, where n is an integer that represents the number of Gaussian primitive functions in each basis function.

In *split-valence basis sets*[46], each valence atomic orbital is described by *n* basis functions (Double Zeta - DZ, Triple Zeta - TZ,...*n*Z) and each core orbital by a single basis function. For instance, in the 6-31G basis set, each core orbital is described by a contracted Gaussian, which is a linear combination of 6 primitive Gaussians; and each valence orbital consists of two contracted Gaussians of 3 primitive and 1 primitive Gaussians, respectively.

Polarized basis sets[46] include polarization functions with an angular momentum greater than that required for the description of the atomic ground state, in order to correctly describe the changes in the electron density of an atom in a molecular environment. Using the previous example, a 6-31G(d,p) basis set, also expressed as 6-31G**, indicates that *d*-functions have been added to heavy atoms and *p*-functions to hydrogen atoms.

There are also basis sets that include *diffuse* functions. These are *s* and *p* functions, whose spatial extension is much greater than those in the description of the valence orbitals, allowing

electrons to occupy a larger region. It is important to include them for systems with electrons that are far from the nucleus. In the 6-31+G** basis set, a diffuse function has been added on heavy atoms. A second + sign indicates the addition of another diffuse function on the hydrogen atoms.

The basis sets we have talked about so far were introduced by John Pople[47,48]. In these sets of functions, the exponents and the contractions coefficients were optimized without considering the effects of the electronic correlation. However, these functions are very often used in post Hartree-Fock calculations, where such effects are considered. To overcome this contradiction, various basis sets have been built to be used in these methods, including those developed by Dunning and his colleagues, known as *correlation-consistent basis sets*[49]. Examples of them are cc-pVDZ, cc-pVTZ, etc… The addition of diffuse polarized functions gives rise to the augmented series[50] (aug-cc-pVDZ, aug-cc-pVTZ,…), suitable for anions and systems with hydrogen bonds.

In the same line of the correlation-consistent basis sets, we can find *polarization-consistent basis sets*[51] (pc-n). Unlike all the sets mentioned above, theses basis have been specifically designed for density functional theory methods, enabling a quicker convergence. As an example, pc-1 and pc-2 would be the equivalent of cc-pVDZ and cc-pVTZ respectively. The augmented series (aug-pc-n) are also obtained by adding diffuse functions.

2.2.5. Summary

In summary, the accuracy of the results and the computational cost in electronic structure calculations depend on two factors: the level of theory of the method used, which accounts for the electronic correlation, and the size of the basis set that describes the

wavefunction or the electron density. Therefore, finding the best possible combination of the two becomes crucial in arriving at results that are as close as possible to the experiment. This is summarized in Figure 1[52], where we can see a hierarchy of methods across the level of theory and the size of the basis sets that we have previously described.

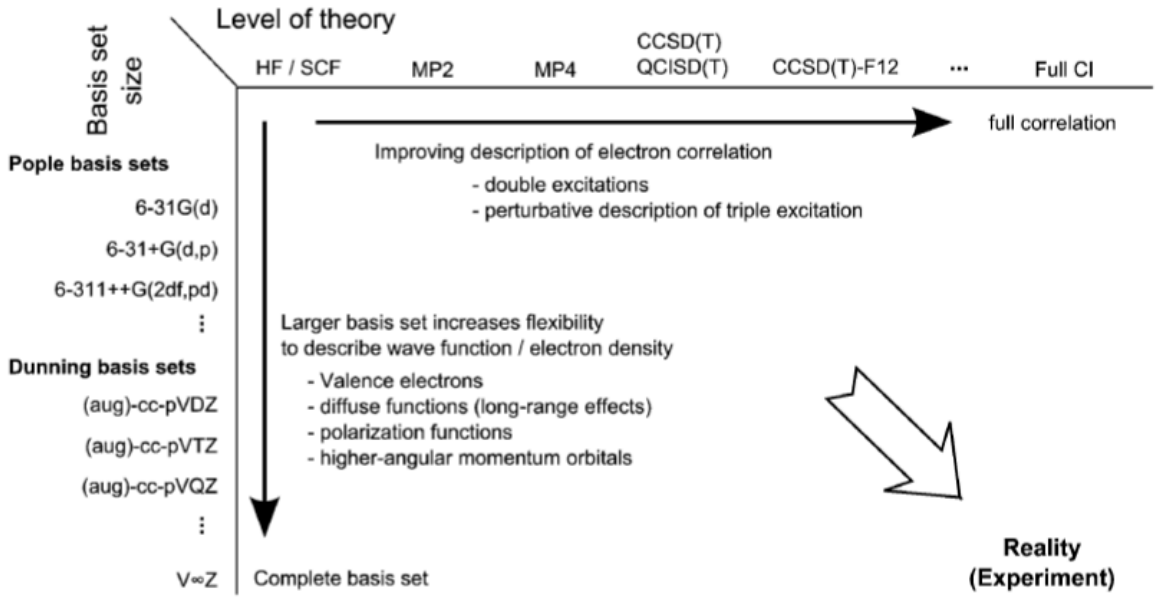


Figure 1. Hierarchy of electronic structure methods and basis sets. Caption taken from Vereecken et al.[52]

2.3. Molecular Mechanics Force Fields

In the context of Molecular Mechanics (MM), molecules are depicted as a combination of spheres (atoms) connected by springs (chemical bonds), which motion can be described with the laws of classical physics[36]. Like in its analogue case in quantum mechanics, the *Born-Oppenheimer approximation* plays an important role since only the motion of the nuclei are considered. Electrons are ignored in the MM models. Therefore, the energy of the molecule can be described as a function of the coordinates of the atomic nuclei. This function is called *Force Field.* Changes in the potential energy of a system can be represented as a surface, called

Potential Energy Surface (PES). One of the aims in molecular modelling is finding the minimum points in the PES that correspond to optimized molecular structures.

2.3.1. Force Field Equations

The term *force field* arises from the fact that the force on a particle is the negative of the first derivative of the potential energy of that particle with respect to displacement along some direction. Therefore, a force field can be differentiated to get the force on each particle. The majority of the force fields used to describe molecular systems can be defined by a sum of *bonded terms* and *nonbonded terms*[34].

$$V = \sum_{\text{bonds}} V_{stretch} + \sum_{\text{angles}} V_{bend} + \sum_{\text{dihedrals}} V_{torsion} + \sum_{\text{pairs}} V_{vdW} + \sum_{\text{pairs}} V_{elec} \qquad (2.3)$$

The bonded terms include energy contributions from bond stretching ($V_{stretch}$), bond angle bending (V_{bend}) and torsional motion around single bonds ($V_{torsion}$). On the other hand, the nonbonded terms include the van der Waals (V_{vdW}) and electrostatic (V_{elec}) interactions, accounting for the interactions between atoms pairs that are not directly bonded together. The main components of a standard force field are described below.

Bond Stretching

Generally, the potential energy for a covalent bond can be described with a Morse function[53]. Although, the *classic harmonic oscillator* approximation is widely accepted in MM force fields since it is good enough for near-equilibrium situations and more efficient in terms of computational cost[28].

$$V_{stretch} = k_{stretch} \left(l - l_{eq} \right)^2 \qquad (2.4)$$

where $k_{stretch}$ is a proportionality constant (one-half of the bond *force constant*), l is the length of the bond when stretched and l_{eq} is the length of the bond in equilibrium

Angle bending

Equally, the deviation of angles from their reference values can be approximated by a harmonic potential.

$$V_{bend} = k_{bend}\left(\theta - \theta_{eq}\right)^2 \tag{2.5}$$

where k_{bend} is also a proportionality constant (one-half of the *angle bending force constant*), θ is the size of the bond angle when distorted and θ_{eq} is the size of the bond angle in equilibrium.

Torsional Term

The torsional term describes the change in energy associated with the rotation around a B-C bond in a series of four atoms A-B-C-D, where A-B, B-C and C-D are bonded. The potential is characterized by a 360º periodicity and it is commonly expressed as a cosine series expansion.

$$V_{torsion} = \sum_n \frac{k_{rot}}{2}\left[\, 1 + \cos\left(n\varphi + \varphi_n\right)\right] \tag{2.6}$$

where k_{rot} is the *torsional rotation force constant* (the height of the torsion barrier around the B-C bond), n is the multiplicity (number of minima in the function when rotating 360º), φ is the torsion angle and φ_n the phase angle (usually chosen to have an energy minimum at 180º).

Van der Waals Interaction

Van der Waals interactions[28] arise from instantaneous non-bonding interactions between the electronic densities. When two nonbonded atoms get closer, there is a moment when their

electronic clouds begin to overlap. That distance is the sum of the van der Waals radius of both atoms. Up until that point, the electronic motion of both atoms is correlated and they develop induced electric dipoles that are mutually attractive. This interaction is called *London dispersion* or *van der Waals dispersion,* and it is inversely proportional to the sixth power of the distance between atoms.

If both atoms continue to get closer to a distance that is shorter than the sum of their van der Waals radii and do not form a bond, another repulsion interaction arises due to the electronic density cloud overlapping. This interaction, known as *Pauli repulsion interaction* or *exchange interaction*[54]*,* increases very rapidly as the distance becomes smaller and contributes to the nonbonded interaction energy too.

Both terms, the London dispersion and the Pauli repulsion interaction, are included in a single equation, often referred to as the *Lennard-Jones potential*[55]:

$$V_{vdW}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.7}$$

For each pair of atoms, ε_{ij} represents the depth of the potential well, σ_{ij} the collision diameter (the distance at which the inter-particle potential is zero) and r_{ij} the distance between the atoms.

Electrostatic Interaction

The electrostatic interactions between atom pairs are described using the Coulomb potential classical formula:

$$V_{elec}(r_{ij}) = \frac{q_i\, q_j}{4\pi\, \varepsilon_0\, r_{ij}} \tag{2.8}$$

where ε_0 represents the vacuum permittivity, q_i and q_j the atomic charges of each atom, and r_{ij} the distance between atoms.

2.3.2. Force Field Parameterizations

In molecular mechanics, parameters refer to the values that force constants and equilibrium geometries take in a potential energy function. The quality and the precision of the results of a force field depend on the functions that describe the energy as well as its parameters. These parameters are often created on quantum mechanical calculations of small model compounds and experimental data obtained from x-ray crystallography, NMR, spectroscopy, inelastic neutron scattering and thermodynamics[36].

It is important to obtain or generate an appropriately parameterised force field for each system under investigation and this process is complicated and labour-intensive. Besides, functional forms and parameters are developed for a particular force field and cannot reliably be mixed[56]. Although the majority of force fields have been devised and optimized for many common classes of molecules, such as biomolecules (protein, lipid, carbohydrate and nucleic acid) and simple organic molecules, there is usually a trade-off to be made between the accuracy of any given force field parameterization and its range of applicability. The more generally applicable a force field parameter set is, the less accurate it is likely to be. An example of this is the Universal Force Field (UFF)[57], which can be applied to a wide range of systems although it is generally not very accurate for any of them. Some of the most commonly used force fields for biomolecules are the CHARMM[58], GROMOS[59] and AMBER[60] force fields.

2.4. Quantum Nuclear Vibrational Models

In the first half of this chapter, we have seen how potential energy surfaces are constructed from quantum and classical points of view. Once a PES is available, the nuclear vibrational Schrodinger equation can be solved to find the nuclear vibrational wavefunctions and vibrational modes of the system.

In order to understand how nuclear vibrations are modelled in a quantum mechanics framework, we will first use the simplest model, a diatomic molecule. Afterwards, we will establish a parallelism between diatomic and polyatomic molecules that will help us better understand the relation between the structure, energetics and vibrations in them.

2.4.1. Diatomics

A diatomic molecule is a molecule composed of two atoms linked by a chemical bond. The energy required to stretch a chemical bond can be described by a potential function. An example of it is shown in Figure 2, where the red curve represents the potential energy V(x), known as Morse potential[53], as a function of the displacement *x* from the equilibrium bond length for a diatomic molecule. The zero of energy is chosen to be the bottom of the potential, but the energy that the bond has in the equilibrium is not zero as we will see later. The asymptotic behaviour for large values of *x* would give us the disassociation energy of the bond, i.e. the energy required to separate both atoms[37].

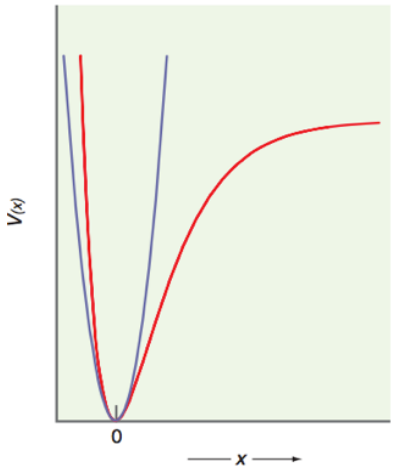


Figure 2. Schematic representation of the harmonic oscillator potential and the Morse potential.[37]

For the majority of diatomic molecules, only the two lowest vibrational levels are occupied at room temperature. This fact will allow us to approximate the potential energy red curve, representing the Morse Potential, to the grey one, which represents the harmonic potential. It is visible that for low energies near the equilibrium bond length, both curves are almost the same. The harmonic potential is deduced from the Hooke's law,

$$V(x) = \frac{1}{2}\, kx^2 \qquad (2.9)$$

In the previous equation, k is the force constant, a measure of the bond strength.

Atoms in a molecule are not static, they are constantly vibrating from their equilibrium position (bond length equilibrium). In fact, by using the harmonic oscillator approximation, we are thinking of those chemical bonds as springs between the vibrating atoms like in the classical analogy. But in this case, we try to solve a quantum mechanical problem, not a classical one. Therefore, we need to solve the Schrödinger equation for a system of mass μ ($\mu = \frac{m_1 m_2}{m_1 + m_2}$, where m_1 and m_2 are the atomic mass of atoms 1 and 2, respectively), which states are described by the nuclear wave function, $\psi_n(x)$, to find out the allowed vibrational energies and frequencies of vibration.

$$-\frac{\hbar}{2\mu}\frac{d^2\psi_n(x)}{dx^2} + \frac{kx^2}{2}\psi_n(x) = E_n\psi_n(x) \tag{2.10}$$

The solutions for the energy states are obtained after applying the boundary conditions,

$$E_n = h\nu\left(n + \frac{1}{2}\right) \tag{2.11}$$

$$n = 0, 1, 2, 3, \ldots$$

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{2.12}$$

where n is a quantum number that represents the state the molecule is at. For each energy state, we can obtain the allowed frequency of vibration of the molecule. Therefore, we establish a relation between the energy of the molecule, its frequency of vibration and its bond force constant. Note that for the state n = 0, the energy of the molecule is not zero. That energy is known as the *zero-point vibrational energy*.

But, this is just a harmonic approximation and only gives approximate solutions for the first states of excitation. A more realistic model is based on the Morse potential, which more accurately represents how the energy of the molecule changes as a function of the bond length:

$$V(x) = D_e\left(1 - e^{-\alpha(x - x_e)}\right); \quad \alpha = \sqrt{\frac{k}{2D_e}} \tag{2.13}$$

where D_e is the bond disassociation energy, x_e is the equilibrium bond length and α is a constant that characterizes the shape of the potential and determines the vibrational frequencies. The force constant, like in the harmonic oscillator case, is defined by:

$$k = \left(\frac{d^2V(x)}{dx^2}\right)_{x=x_e} \tag{2.14}$$

Fortunately, we can still find the solutions of the Schrödinger equation for the Morse potential. The energies are:

$$E_n = h\nu\left(n+\frac{1}{2}\right) - \frac{(h\nu)^2}{4D_e}\left(n+\frac{1}{2}\right)^2 \qquad (2.15)$$

The second term is the anharmonic correction to the energy levels found in the harmonic oscillator solutions. The consequence of this correction is non-equidistant energy levels in the molecule. An example of a Morse potential for the HCl molecule is shown in Figure 3. Again, the red line represents the Morse potential as a function of the bond length x, whereas the purple one represents the harmonic potential and the grey horizontal lines the allowed energy levels for this molecule. D_0 is the bond energy defined with respect to the lower state, D_e is the disassociation energy and x_e the equilibrium bond length.

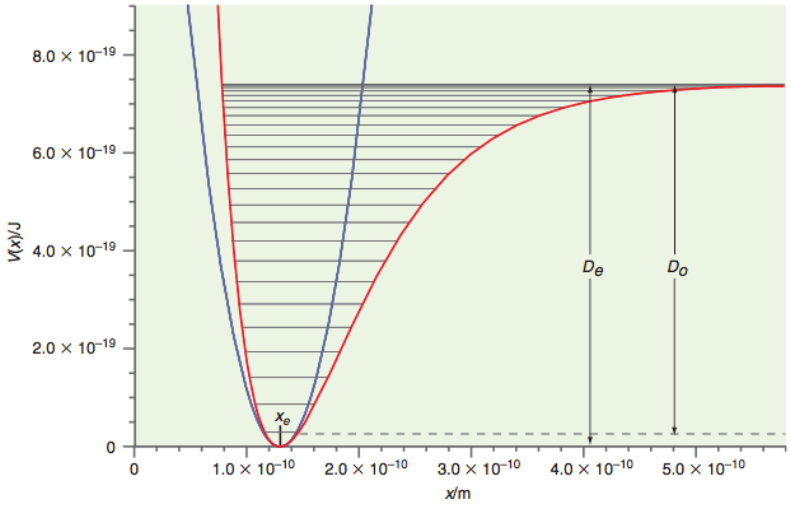


Figure 3. Morse potential for HCl in red, harmonic approximation in purple, anharmonic energy levels in grey [37]

Now, we have a better overview of the behaviour of a chemical bond and how we can relate the force constant to the energies and the vibrational frequencies for a diatomic molecule. The next step would be measuring those frequencies, and we can do it thanks to infrared spectroscopy. This experimental technique is based upon the fact that molecules absorb

radiation which frequency matches the difference between their vibration energy levels. For the diatomic case, the light has to have the right energy to induce 0→1 vibrational transition.

2.4.2. Polyatomics

A polyatomic molecule has more than two atoms connected by chemical bonds, therefore, the problem cannot be seen as a one-dimensional interaction as in a diatomic molecule. We will illustrate parallels between the study of the structure, energy and vibrations of diatomic and polyatomic molecules in Table 2.

	Diatomic molecule	**Polyatomic molecule**
Structure	Equilibrium bond length	Optimized geometry
Energy	Potential energy curve	Potential energy surface (PES)
Vibrations	Quantum harmonic oscillator	Normal mode analysis: - Uncoupled modes of motion - Energy changes harmonically
	Quantum Morse oscillator (anharmonic)	Nuclear vibrational wave function methods
IR Spectroscopy	Transitions between 0→1 quantum states	Transitions between 0→1 quantum states for each vibrational mode

Table 2. Parallel between models for determining the structure, energy and quantum vibrational states of diatomic and polyatomic molecules

Structure. In a diatomic molecule, we only need to know the length of the bond between the two atoms. In the case of a polyatomic, we need to know the molecular geometry, which includes bond length, bond angles and torsion angles between the atoms in the molecule. Experimentally, it can be determined by spectroscopic and diffraction methods, although it can be also computed by *ab initio* computational methods.

Energy. As we have already seen, once we know the structure of the molecule, we can create a potential energy surface (PES), which is the translation into (3N-6) dimensions of the potential energy curve for a diatomic molecule (3N-6 are the degrees of freedom for general molecules, where N is the number of atoms in the molecule). A potential energy surface is used to describe the energy of the system in terms of the position of the atoms. In other words, it describes energy as a function of atomic displacement and bond bending angle. In Figure 4, we present the potential energy surface of a molecule of water as an example. P_{min} corresponds to the point of equilibrium of the water molecule, i.e. the minimum energy geometry for the three atoms (Note that in this diagram only 2 of the 3 coordinates are illustrated).

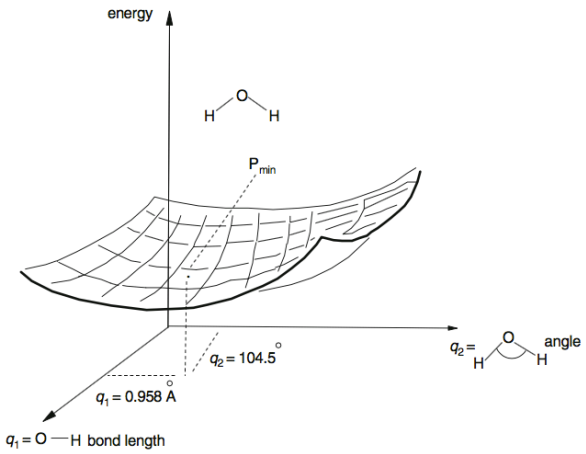


Figure 4. H_2O potential energy surface[1].

In order to calculate potential energy surfaces in the context of quantum mechanics, we need to use the computational chemistry methods that were previously described in the electronic structure methods section.

Vibrations. Like in the diatomic molecule case, atoms in a polyatomic molecule vibrate. In fact, we can use the classic harmonic oscillator, as a first approximation, to find the relation between the bond strengths and the frequencies of vibration. Therefore, we need to apply *normal mode analysis* (NMA)[61], in which systems of atoms are treated as a collection of small displacements about an equilibrium position, oscillating with characteristic angular frequencies, where the net force on all the atoms is zero. In this approximation, we define a linear combination of those atomic displacements in a way that the energy changes in the most uncoupled way, i.e. as independently as possible. In order to do that, we need to make two assumptions. The first one is that the energy can be separated into modes and, the second one, that along each mode, the energy is harmonic[62].

$$\Delta E_{total} = \sum_{\text{modes}} \Delta E_i \,; \quad i = 1, 2, 3, \ldots \tag{2.16}$$

$$\Delta E_i = k \left(\mathrm{Q} - \mathrm{Q_i} \right)^2 \tag{2.17}$$

Where ΔE_i is the harmonic energy along each mode and $\left(\mathrm{Q} - \mathrm{Q_i} \right)$ is the set of linear combinations of atomic displacements that ensure that the energy changes independently along each mode, to second order.

The next step, then, would be considering a more realistic model where anharmonicity and mode coupling are considered, i.e. including terms higher than second order in the potential energy surface expansion and also incorporating terms that involve concurrent displacement along multiple modes concurrently. However, this means that the nuclear vibrational Schrodinger equation can no longer be solved exactly simply by ensuring that the energy changes in a decoupled way along vibrational models.

Like in the case of the electronic structure methods, there is a hierarchy of models for solving this problem, all starting from a simpler exactly solvable harmonic model and incorporating

anharmonic corrections either via perturbation theory or configuration interaction type approaches[63]. In these methods, we can include *vibrational self-consistent field* (VSCF), *vibrational configuration interaction* (VCI), *vibrational Møller-Plesset theory* (VMP), and *vibrational coupled cluster theory* (VCC)[63]. As observed for analogous electronic structure methods, going higher in the hierarchy implies more accuracy in the expected results, but also an increase in the computing time. Therefore, it is important to find a balance between the two factors. We show this hierarchy in Table 3, increasing in complexity from top to bottom.

Nuclear vibration methods
Normal mode analysis (NMA)
Self-consistent field theory (VSCF)
Vibrational perturbation theory (VPT)
Vibrational configuration interaction (VCI)
Vibrational coupled cluster theory (VCC)

Table 3. Nuclear vibration models, ordered by increasing accuracy and computational cost.

2.5. Molecular Dynamics

Molecular dynamics[1] (MD) is a computational model that simulates the behaviour or evolution of a system over time, calculating the forces between the particles that make it up and solving Newton's equations of motion. This technique allows calculating the trajectories of the atoms or particles that form a macroscopic system at different instants of time, obtaining a set of positions and speeds that correspond to a sampling of the phase space of the system,

thus simulating its microscopic behaviour. From this information and using Statistical Mechanics, we can obtain macroscopic thermodynamic properties of the system.

It is a deterministic method so, by knowing an initial configuration of the system, that is, the positions **r,** and speeds **v**, of each of the particles that compose it, it is possible to predict the positions and speeds at a future moment by solving Newton's equations of motion. For a system of N particles, Newton's equations are given by:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i ; \qquad i = 1, 2, \ldots, N \tag{2.18}$$

$$\mathbf{v}_i = \frac{d\mathbf{r}_i}{dt} \tag{2.19}$$

where $\boldsymbol{F}_i$ is the total force acting on the i particle. If the forces acting on the i particle are conservative, they can be calculated based on the gradient of the potential energy or *force field*, V:

$$\mathbf{F}_i = -\nabla V \tag{2.20}$$

2.5.1. Numerical Integrators

Newton's equations of motion for more than two particles do not have analytical resolution, but they can be solved numerically. To do this, it is first necessary to transform the continuous temporal evolution of the system into a discretized evolution, characterized by Δt, a *time step* that occurs between one microscopic state of the system and the next. There are various algorithms for solving these equations. When choosing an algorithm for the numerical resolution of these equations, it is important to take into account both the accuracy of the results

it provides and the efficiency and computational performance. Among the most widely used is the *leap-frog algorithm* due to its simplicity, efficiency and stability[64].

The leap-frog algorithm uses the positions **r** of an instant of time t and the velocities **v** of a previous instant of time, $(t - 1/2\ \Delta t)$, to obtain the velocities at a future instant of time, $(t + 1/2\ \Delta t)$, and finally, obtain the positions at the instant of time $(t + \Delta t)$ using the following expressions:

$$\mathbf{v}(t + 1/2\ \Delta t) = \mathbf{v}(t - 1/2\ \Delta t) + \frac{\Delta t}{m}\mathbf{F}(t) \tag{2.21}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t\mathbf{v}(t + 1/2\ \Delta t) \tag{2.22}$$

where $\mathbf{F}(t)$ are the forces acting on the system at a time instant t. During the passage of time, those forces are considered to remain constant. The time-step, Δt, must be small enough to correctly sample the evolution of the system, normally being on the order of femtoseconds.

Once the velocities and positions of the new state of the system have been calculated by solving Newton's equations, the forces acting on each of the particles are reassessed and a new cycle of molecular dynamics begins. By repeating this process, the time evolution of the system is obtained.

2.5.2. Molecular Dynamics Simulations

Three stages can be distinguished in a Molecular Dynamics simulation:

- *Thermalization:* Once the initial coordinates have been assigned to each particle in the system, the simulation begins by randomly assigning initial speeds to each particle

following a Boltzmann distribution consistent with the required temperature given by the theorem of equipartition of energy. The initial positions of the atoms can be obtained from experimental structures, determined by X-ray or NMR, or modelled theoretically.

Initially, the velocities of all the particles are rescaled to achieve a zero or resulting net linear momentum. During thermalization, a simulation is carried out in which the speeds are rescaled repeatedly to bring the system to the required temperature and dynamic equilibrium.

When speed rescaling is no longer necessary, the temperature has stabilized and the thermal equilibrium state has been reached, terminating the thermalization stage.

- *Equilibration*: A given period of time, which contains a significant number of simulation steps where it is not necessary to rescale the velocities, and which ensures that the Hamiltonian of the system remains constant and that the system is in equilibrium.

- *Production*: Once equilibrium is reached, the system is allowed to evolve and positions and velocities of the particles over time are stored for later analysis. The information obtained in this stage will be used to calculate the static and dynamic properties of the system.

2.5.3. Ensembles in Molecular Dynamics

As we have already said, by solving Newton's equations of motion, the trajectories and velocities of each molecule that make up the system are generated at different instants of time. If the system is closed and isolated from the outside by rigid walls, the resolution of these

equations generates a distribution of microstates corresponding to a macrostate whose energy remains constant. That is, the simulation of a system whose number of particles (N), volume (V) and energy (E) remains constant over time or, in other words, would be carrying out the simulation of a system in the microcanonical ensemble (NVE). However, most experiments with real substances do not occur under these conditions.

In certain situations, instead of setting the total energy of the system, it is much more interesting and practical to set the temperature. To carry out a simulation at constant temperature (as in the NVT and NPT ensembles), the use of a thermostat is necessary. The use of a thermostat in MD simulations allow setting the temperature of the system, which is very useful when you want to study the behaviour of a system at a certain temperature. Among the most widely used thermostats are:

- *Berendsen Thermostat*[65]. In this thermostat, the system is coupled to an external thermal bath whose temperature is fixed at the temperature of interest. In this case, the speeds of each particle will be rescaled to adjust the system temperature. A drawback of this thermostat is that, formally, the generated collective is not strictly a canonical ensemble because it does not allow the kinetic energy to fluctuate.

- *V-scale Thermostat*[66]. This thermostat is essentially the same as the Berendsen thermostat plus a stochastic term that ensures a correct distribution of kinetic energy.

- *Nosé-Hoover Thermostat*[67,68]. The Nosé-Hoover algorithm modifies Newton's equations by introducing a new term, with a friction coefficient that adjust the velocity of the particles of the system in order to maintain the desired temperature. The equations of motion modified by this thermostat are reversible in time and

deterministic. However, the temperature fluctuations it generates can become almost periodic.

Just as it is interesting to set the temperature of a system, it can be equally interesting to set the pressure. In the isothermal-isobaric ensemble or NPT, in addition to setting the number of molecules, temperature and pressure are also set. To set the pressure throughout a simulation, the use of a barostat will be necessary. Among the most used barostats we can find:

- *Berendsen Barostat*[69]. Berendsen's algorithm rescales the coordinates and vectors that form the simulation box at each simulation step using a rescaling matrix. In this procedure, the speeds are neither rescaled nor rotated due to the change in size and shape of the simulation box and it cannot be ensured that the NPT ensemble is being correctly sampled.

- *Parrinello-Rahman Barostat*[70,71]. Similarly to the Nosé-Hoover thermostat, the Parrinello-Rahman barostat adjusts the system pressure by correctly sampling the ensemble NPT. Again, this barostat rescales the coordinates and vectors that make up the simulation box at each simulation step using a rescaling matrix. By using this barostat the equations of motion are also modified.

2.5.4. Periodic Boundary Conditions

Molecular Dynamics simulations aim to study and predict the macroscopic behaviour of a real system from a microscopic point of view. In order to keep the system in a condensed phase and prevent the molecules that compose it from moving away from each other, it is necessary to establish a potential that acts in the same way as a container would (boundary conditions), preventing the system from expanding beyond the its walls. Given the finite size

of the simulated systems, the use of a "potential wall" is not a feasible solution due to the effects these boundary conditions have on the properties of the system. A molecule will not have the same behaviour in a homogeneous phase as in contact with a "potential wall".

In order to avoid this problem and to be able to simulate a bulk phase without its properties being affected by the proximity of the molecules that make up the system to a potential wall, it is necessary to use *periodic boundary conditions*[72,73]. To do this, the system is replicated in space, forming an infinite network of identical systems (Figure 5), so that now the system of interest is no longer confined by walls, but is surrounded by identical systems, that make it "feel" like it's inside a bulk phase. Since all the replicated systems are identical, if during the course of the simulation, any of the molecules that make up the original system leave it, a molecule identical to the previous one will re-enter the system from one of the adjacent replicated systems. Therefore, the density and the number of particles in the system remain constant, without the need for confining walls and thus avoiding the effect of the contour surfaces on the properties of the system. During the simulation, it is not necessary to store the coordinates of each of the molecules in all the replicated systems, but only those of the original system.

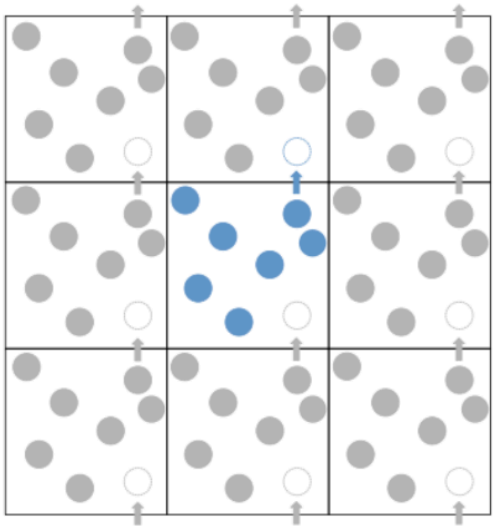


Figure 5. Two-dimensional graphical representation of periodic boundary conditions.

As previously stated, Molecular Dynamics studies the temporal evolution of a system by calculating the forces acting on each of the molecules that make up the system. This force is calculated from the contributions of potential energy in which each molecule is involved. If we consider an additive potential in pairs, such as the Lennard-Jones potential or the Coulomb potential, the total force undergone by particle *i* will be calculated as the $N - 1$ contributions due to the interaction of all the particles of the system with particle *i*. Furthermore, it would also be necessary to take into account the interaction of particle *i* with all the particles of the replicated systems. This supposes an infinite number of interactions, something impossible to calculate in practice. For this, the *minimum image criterion*[72,73] is established, which implies that the particle *i* only interacts with the closest periodic images of the other N – 1 particles that make up the system.

In order to save computation time, it is possible to define a *cut-off radius*[72,73] for the interaction potential, such that the interaction between those particles located at a distance greater than the cut-off radius, whether they belong to the system original or its minimal images, are not taken into account.

2.6. Applications

This thesis aims to explain the vibrational behavior of a wide range of different chemical systems of varying size and complexity: from small gas phase molecules, through to the behavior of isolated or semi-isolated clusters within crystalline environments, all the way up to modelling the spectroscopic properties of bulk cellulose and their temperature-dependence. Therefore, it is necessary to carefully select approaches that strike the optimal balance between accuracy and computational cost for each system of interest.

In the first results chapter, we will explore different ways of computing *ab initio* PES and solving the quantum nuclear vibrational Schrödinger equation, and specifically looking at the relationship between level of *ab initio* theory used in constructing the PES and the accuracy of the predicted results. Small gas phase molecules are deliberately used as test systems, so that all models can be assessed for accuracy without regard to computational cost.

The second and third results chapters focus on modelling the experimentally-observed IR spectra of dichloride hydrates that are trapped within crystalline environments. Because these systems are much larger, making quantitative predictions of the peak positions is much harder, and we therefore aim to qualitatively interpret the observed spectra, assigning the character of the vibrational modes using empirically-corrected harmonic frequency calculations.

Finally, the aim of the last results chapter is to explain how hydrogen bond stretching frequencies change with temperature within crystalline cellulose. Due to the size of the system and the fact that quantum nuclear vibrational models do not capture thermal effects, the most suitable approach is molecular dynamics.[74] Furthermore, high quality parameterizations make force fields the most sensible choice for describing the cellulose PES.[75]

References

(1) Lewars, E. G.: Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics; Springer International Publishing, **2016**.

(2) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biology* **2011**, *9*, 71.

(3) Young, D. C.; Nelson Cpa, M. M. S. S. L.: *Computational Drug Design : A Guide for Computational and Medicinal Chemists*; John Wiley & Sons, **2009**.

(4) De Benedetti, P. G.; Fanelli, F. Computational modeling approaches to quantitative structure–binding kinetics relationships in drug discovery. *Drug Discovery Today* **2018**, *23*, 1396-1406.

(5) Ong, S. P.; Wang, L.; Kang, B.; Ceder, G. Li−Fe−P−O2 Phase Diagram from First Principles Calculations. *Chemistry of Materials* **2008**, *20*, 1798-1807.

(6) Li, L.; Muckerman, J. T.; Hybertsen, M. S.; Allen, P. B. Phase diagram, structure, and electronic properties of $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solutions from DFT-based simulations. *Physical Review B* **2011**, *83*, 134202.

(7) Jithesh, K.; Govind; Waghmare, U. V.; Shivaprasad, S. M. Experimental deduction of In/Si(111) 2D phase diagram and ab initio DFT modeling of 2√3 phase. *Applied Surface Science* **2009**, *256*, 348-352.

(8) Adjaoud, O.; Steinle-Neumann, G.; Burton, B. P.; van de Walle, A. First-principles phase diagram calculations for the HfC--TiC, ZrC--TiC, and HfC--ZrC solid solutions. *Physical Review B* **2009**, *80*, 134112.

(9) Rigo, V. A.; Miranda, C. R.; Baletto, F. Ethanol chemisorption on core–shell Pt-nanoparticles: an ab initio study. *The European Physical Journal B* **2019**, *92*, 24.

(10) Xu, C.-Q.; Lee, M.-S.; Wang, Y.-G.; Cantu, D. C.; Li, J.; Glezakou, V.-A.; Rousseau, R. Structural Rearrangement of Au–Pd Nanoparticles under Reaction Conditions: An ab Initio Molecular Dynamics Study. *ACS Nano* **2017**, *11*, 1649-1658.

(11) Kilin, D. S.; Prezhdo, O. V.; Xia, Y. Shape-controlled synthesis of silver nanoparticles: Ab initio study of preferential surface coordination with citric acid. *Chemical Physics Letters* **2008**, *458*, 113-116.

(12) Munejiri, S.; Shimojo, F.; Hoshino, K.; Inui, M. Structural studies on fluid sulfur at high temperatures and high pressures: II. Molecular structure obtained by ab initio molecular dynamics simulations. *Journal of Non-Crystalline Solids* **2019**, *510*, 15-19.

(13) Lapinski, L.; Nowak, M. J.; Les, A.; Adamowicz, L. Ab Initio Calculations of IR Spectra in Identification of Products of Matrix Isolation Photochemistry: Dewar Form of 4(3H)-Pyrimidinone. *Journal of the American Chemical Society* **1994**, *116*, 1461-1467.

(14) Wang, Y.; Carter, S.; Bowman, J. M. Variational Calculations of Vibrational Energies and IR Spectra of trans- and cis-HOCO Using New ab Initio Potential Energy and Dipole Moment Surfaces. *The Journal of Physical Chemistry A* **2013**, *117*, 9343-9352.

(15) Dargelos, A.; Karamanis, P.; Pouchan, C. Ab-initio calculations of the IR spectra of dicyanodiacetylene (C6N2) beyond the harmonic approximation. *Chemical Physics Letters* **2019**, *723*, 155-159.

(16) Escribano, R.; Gómez, P. C.; Maté, B.; Molpeceres, G.; Artacho, E. Prediction of the near-IR spectra of ices by ab initio molecular dynamics. *Physical Chemistry Chemical Physics* **2019**, *21*, 9433-9440.

(17) Born, M.; Oppenheimer, R. Zur quantentheorie der molekeln. *Annalen der Physik* **1927**, *389*, 457-484.

(18) Tsutsumi, T.; Harabuchi, Y.; Yamamoto, R.; Maeda, S.; Taketsugu, T. On-the-fly molecular dynamics study of the excited-state branching reaction of α-methyl-cis-stilbene. *Chemical Physics* **2018**, *515*, 564-571.

(19) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701-1718.

(20) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, *91*, 1-41.

(21) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **2011**, *27*, 1575-1577.

(22) Szabo, A.; Ostlund, N. S.: *Modern quantum chemistry: introduction to advanced electronic structure theory*; Dover Publications: Mineola, N.Y, **1996**.

(23) Helgaker, T.; Jørgensen, P.; Olsen, J.: *Molecular electronic-structure theory*; Wiley, **2012**.

(24) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society* **1928**, *24*, 89-110.

(25) Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik* **1930**, *61*, 126-148.

(26) Löwdin, P.-O. Quantum theory of many-particle systems. III. Extension of the Hartree-Fock scheme to include degenerate systems and correlation effects. *Physical Review* **1955**, *97*, 1509.

(27) Sherrill, C. D.; Schaefer III, H. F.: The configuration interaction method: Advances in highly correlated approaches. In *Advances in quantum chemistry*; Elsevier, **1999**; Vol. 34; pp 143-269.

(28) Cramer, C. J.: *Essentials of computational chemistry: theories and models*; John Wiley & Sons, **2013**.

(29) Schrödinger, E. Quantisierung als Eigenwertproblem. *Annalen der Physik* **1926**, *385*, 437-490.

(30) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, *46*, 618-622.

(31) Hattig, C. Beyond Hartree-Fock: MP2 and coupled-cluster methods for large systems. *Computational Nanoscience: Do It Yourself* **2006**, *31*, 245-278.

(32) Handy, N.; Knowles, P.; Somasundram, K. On the convergence of the Møller-Plesset perturbation series. *Theoretica Chimica Acta* **1985**, *68*, 87-100.

(33) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864-B871.

(34) Levine, I. N.: *Quantum Chemistry*; Pearson, **2014**.

(35) Bene, E.; Nagy, A. The correlation energy in terms of density moments along the adiabatic connection in the density functional theory. *Chemical Physics Letters* **2000**, *324*, 475-481.

(36) Ramachandran, K.; Deepa, G.; Namboori, K.: *Computational chemistry and molecular modeling: principles and applications*; Springer Science & Business Media, **2008**.

(37) Engel, T.: *Quantum chemistry and spectroscopy*; Pearson Higher Ed, **2012**.

(38) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **1965**, *140*, A1133.

(39) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **1980**, *58*, 1200-1211.

(40) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, *38*, 3098.

(41) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **1992**, *45*, 13244.

(42) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, *37*, 785.

(43) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics* **1993**, *98*, 1372-1377.

(44) Ulusoy, I. S.; Wilson, A. K.: Slater and Gaussian basis functions and computation of molecular integrals. In *Mathematical Physics in Theoretical Chemistry*; Elsevier, **2019**; pp 31-61.

(45) March, N.; Nesbet, R. Self-Consistent Fields in Atoms. *Physics Today* **1976**, *29*, 50.

(46) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *The Journal of Chemical Physics* **1971**, *54*, 724-728.

(47) Wiberg, K. B. Ab Initio Molecular Orbital Theory by W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople. *Journal of Computational Chemistry* **1986**, *7*, 379-379.

(48) Davidson, E. R.; Feller, D. Basis set selection for molecular calculations. *Chemical Reviews* **1986**, *86*, 681-696.

(49) Jr., T. H. D. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007-1023.

(50) Kendall, R. A.; Jr., T. H. D.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992**, *96*, 6796-6806.

(51) Jensen, F. Polarization consistent basis sets: Principles. *The Journal of Chemical Physics* **2001**, *115*, 9113-9125.

(52) Vereecken, L.; Francisco, J. S. Theoretical studies of atmospheric reaction mechanisms in the troposphere. *Chemical Society Reviews* **2012**, *41*, 6259-6293.

(53) Morse, P. M. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Physical Review* **1929**, *34*, 57-64.

(54) Griffiths, D. J.; Schroeter, D. F.: *Introduction to quantum mechanics*; Cambridge University Press, **2018**.

(55) Jones, J. E. On the determination of molecular fields.—II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1924**, *106*, 463-477.

(56) Monticelli, L.; Tieleman, D. Force Fields for Classical Molecular Dynamics. *Methods in Molecular Biology (Clifton, N.J.)* **2013**, *924*, 197-213.

(57) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024-10035.

(58) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **1983**, *4*, 187-217.

(59) van Gunsteren, W. F.; Billeter, S.; Eising, A.; Hünenberger, P.; Krüger, P.; Mark, A.; Scott, W.; Tironi, I. Biomolecular simulation: the GROMOS96 manual and user guide. *Vdf Hochschulverlag AG an der ETH Zürich, Zürich* **1996**, *86*.

(60) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157-1174.

(61) Atkins, P. W.; De Paula, J.; Keeler, J.: *Atkins' physical chemistry*; Oxford University Press, **2018**.

(62) Dykeman, E. C.; Sankey, O. F. Normal mode analysis and applications in biological physics. *Journal of Physics: Condensed Matter* **2010**, *22*, 423202.

(63) Christiansen, O. Vibrational structure theory: New vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties. *Physical Chemistry Chemical Physics* **2007**, *9*, 2942-2953.

(64) Berendsen, H. J.; Van Gunsteren, W. F. Practical algorithms for dynamic simulations. *Molecular-dynamics Simulation of Statistical-mechanical Systems* **1986**, 43-65.

(65) Berendsen, H.: Transport properties computed by linear response through weak coupling to a bath. In *Computer Simulation in Materials Science*; Springer, **1991**; pp 139-155.

(66) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*, 014101.

(67) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **1984**, *52*, 255-268.

(68) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* 1985, *31*, **1695**.

(69) Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684-3690.

(70) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182-7190.

(71) Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **1983**, *50*, 1055-1076.

(72) Ciccotti, G.; Ferrario, M.; Schuette, C. Molecular dynamics simulation. *Entropy* **2014**, *16*, 1.

(73) Allen, M. P. Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins* **2004**, *23*, 1-28.

(74) Roy, T. K.; Gerber, R. B. Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications. *Physical Chemistry Chemical Physics* **2013**, *15*, 9468-9492.

(75) Hansen, H. S.; Hünenberger, P. H. A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. *Journal of Computational Chemistry* **2011**, *32*, 998-1032.

Chapter 3

Benchmarking Potential Energy Surface Construction Schemes

"If you know the enemy and know yourself, you need not fear the results of a hundred battles."

Sun Tzu

3.1. Introduction

Over the last three decades, increasing computer power and developments in algorithms for solving the nuclear vibrational Schrödinger equation have made it possible to routinely and accurately predict and assign vibrational spectra for small to medium sized molecules.[1] The computational bottleneck that prevents these approaches being extended to larger molecules is the need to first construct multidimensional anharmonic potential energy surfaces (PES),[2] which describe how the electronic energy changes as a function of molecular conformation. This is a difficult problem due to what is often referred to as the "curse of dimensionality" – the energy of the system changes as all atoms in the system move in all directions and *ab initio* electronic structure calculations must be performed to evaluate the energy in all possible geometries - this process to obtain the necessary data to construct the PES function can be very computationally expensive depending on the size of the system and the choice of electronic structure method.

A common approach is to represent the PES as a Taylor series expansion in normal mode coordinates about the global minimum point. This method allows the PES to be separated into harmonic and anharmonic contributions and the associated force constants computed at different levels of theory. In the process of calculating the anharmonic force constants, derivatives of the energy must be evaluated with respect to displacement in normal mode coordinates. The number of force constants and calculations that must be performed in order to evaluate them increases substantially as the order of the Taylor expansion increases. Therefore, it is important to find efficient and accurate ways to compute these force constants, especially for larger systems.

There are two main strategies for reducing the computational cost associated with constructing potential energy surfaces – reduce the time taken to evaluate each energy at each displaced geometry or reduce the number of energy evaluations required to describe the energetically-relevant regions of the PES. This work will focus primarily on the first of these considerations.

Nowadays, the number of theories – and computational methods implemented to give effect to them – available for describing the electronic structure of molecules are numerous,[8] including Hartree-Fock (HF), Møller-Plesset (MP) perturbation theory, configuration interaction (CI) and coupled cluster (CC) theory, together with a suite of methods that all come under the umbrella of density functional theories (DFT). The computational cost and the accuracy of the methods can vary widely but are generally correlated – the more computationally intensive the method the more accurate it is likely to be.

The accuracy of electronic structure models is often assessed by comparing the experimental and theoretically computed equilibrium geometries. However, comparing molecular vibrational frequencies may represent a more rigorous way to do so, and gives more information about the quality of the entire potential energy surface in the vicinity of the minimum energy point, rather than just the ability to recover the minimum energy geometry.[9]

In previous studies, a "hybrid" approach has been used for constructing PES in order to accurately model experimental gas phase fundamental frequencies, in which the harmonic and anharmonic force constants within the Taylor series expansion are computed using different electronic structure methods and/or atomic orbital basis sets.[3-7]

In constructing composite force fields, anharmonic force constants are typically computed at a lower level of ab initio theory, and harmonic force constants at a higher level. In practice, this is typically achieved by computing the entire force field at the lower level of theory and then

replacing the harmonic force constants with values obtained at a higher level of theory, without regard to how the character of the normal mode coordinates may change at different levels of theory. These models are motivated by the fact that the major contribution to the accuracy in the calculation of the frequencies are the harmonic terms of the PES.

Once a PES has been created, a hierarchy of methods are available to solve the nuclear vibrational Schrödinger equation and obtain the desired frequencies. The most widely used method due to its low computational cost is normal mode analysis (NMA),[10] although it does not capture the anharmonicity of the systems under study, There are other methods that, in doing so, guarantee more accurate results, including vibrational self-consistent field theory (VSCF),[11] vibrational perturbation theory (VPT),[1] vibrational configuration interaction (VCI),[1] and vibrational coupled cluster theory (VCC).[1] Again, the accuracy of each approach is normally proportional to its computational cost, and the methods above are listed in order of increasing accuracy and computational cost.[12]

The purpose of this chapter is to find the optimal balance between accuracy and computational cost when choosing electronic structure methods for constructing hybrid PES. Benchmark results will be computed for a test set of 19 small molecules using vibrational configuration interaction theory with spectroscopically accurate force fields. These will then be compared to results obtained from force fields computed *ab initio*, at a range of different levels of theory, in order to isolate the errors due to choices of electronic structure models used for computing harmonic and anharmonic force constants.

3.2. Methods

All the potential energy surfaces in this study were computed using the code PyPES,[13] created by Sibaev et al.[14] The potential energy function is given by:

$$V = V_{ref} + \sum_i F_i Q_i + \frac{1}{2!}\sum_{i,j} F_{ij} Q_i Q_j + \frac{1}{3!}\sum_{i,j,k} F_{ijk} Q_i Q_j Q_k + \frac{1}{4!}\sum_{i,j,k,l} F_{ijkl} Q_i\, Q_j Q_l Q_m$$

Where Q are the rectilinear normal mode coordinates and F, the force constants, are the derivatives of the potential with respect to the normal mode coordinates. The global minimum is considered to be the equilibrium geometry of the system and, in this coordinate system, the first order and off-diagonal second order force constants are zero.

Second order force constants can be obtained analytically *ab initio*, by diagonalization of the mass-weighted Hessian matrix. These correspond to the harmonic force constants output directly by quantum chemistry programs. This process also defines the normal mode coordinates. The process for obtaining the anharmonic force constants –the higher order coefficients– up to a 6th order is more complex.

Following the process within PyPES, anharmonic force constants are first computed in rectilinear normal mode coordinates to 4th order either via numerical differentiation of *ab initio* Hessians or analytically from implemented reference potentials. After that, another force field expansion is constructed in curvilinear normal mode coordinates (also to 4th order). Then, relations between force constants from both expansions are analytically obtained, which allow us to estimate the 5th and 6th order rectilinear normal mode force constants from the asymptotic behaviour of the curvilinear normal mode coordinates. No reduction in mode-mode coupling is applied in this work.

The details of the electronic structure models used to construct the implemented reference PES are given in Table 1. In general, reference PES are represented by force fields expanded in carefully hand-crafted symmetry-adapted non-redundant internal coordinates.

Table 1. List of the molecules used in this study together with the electronic structure methods and basis sets used to construct the PES in the reference studies.

Molecule	Electronic Structure Method	Basis Set	Reference
C_2H_4	CCSD(T)	cc-pVQZ	15
C_3H_2	CCSD(T)	cc-pVQZ	16
C_4	CCSD(T)	cc-pCV5Z	17
CH_4	CCSD(T)	cc-pVTZ	18
F_2O	CCSD(T)	cc-pVQZ	19
H_2CO	CCSD(T)	aug-cc-pVQZ	20
H_2O	CCSD(T)*	TQ5+rel+ACPF/QZ*	21
HO_2^+	CCSD(T)*	TQ5+rel+ACPF/QZ*	21
HOCl	CCSD(T)	aug-cc-pV(Q+d)Z	22
HOF	CCSD(T)	cc-pVQZ	19
N_2H_2	CCSD(T)	cc-pVTZ	23,24
NH_2^-	CCSD(T)	aug-cc-pV5Z	25
NH_3	CCSD(T)	aug-cc-pV5Z	26
NH_4^+	CCSD(T)	cc-pVTZ	27
PH_3	CCSD(T)	aug-cc-pV(Q+d)Z	28
SiF_2	CCSD(T)	cc-pVQZ	29
SiH_4	CCSD(T)	cc-pV(Q+d)Z	30
SO_2	CCSD(T)	cc-pVQZ	29
SO_3	CCSD(T)	aug-cc-pVQZ	31

*Corrections to CCSD(T) have been applied such as the inclusion of scalar relativistic effects and core-valence electron correlation as well as higher-order correlation correction. For full explanation see the reference.

In the first part of this work, we construct PESs for 19 molecules using three different electronic structure methods with the same Dunning correlation-consistent polarized valence triple zeta (cc-pVTZ) basis set: Hartree-Fock (HF), Møller-Plesset second-order perturbation theory (MP2) and couple cluster theory (CCSD(T)). In the second part, we used the same force fields created with the three different methods mentioned above but, in this case, replacing the harmonic force constants with those taken from the high-level *ab initio* force fields compiled in the PyPES library. All the equilibrium geometries, as well as all the Hessian calculations within PyPES for each displaced geometry are computed using CFOUR.[32]

In order to obtain the fundamental vibrational frequencies, we used PyVCI, a Python-based software package that includes VCI algorithms for the resolution of the vibrational Schrodinger equation, using the Watson Hamiltonian with Coriolis coupling terms.[12] The wavefunction is constructed from Hartree products of harmonic oscillators basis functions in normal mode coordinates. Each one of these basis functions is defined by a quantum number that specifies its vibrational state. The VCI basis states are generated by specifying the value of the sum of those quantum numbers, which is referred to as the excitation level. In order words, the excitation level (the value of the sum of the quantum vibrational numbers) specifies the allowed vibrational excitations that contribute to each VCI state. In this work, all VCI state included all configurations with a 6^{th}-order excitation level. The computed fundamental frequencies using both the pure and hybrid force fields are compared against the values taken from the PyPES library in which the VCI wavefunction was also expanded to 6^{th} order.

It is important to emphasize the fact that, although the electronic structure method used in the reference material is the same we used in some of our calculations, CCSD(T), the size of the basis sets is larger in most cases. As a consequence, the resultant force fields and computed force constants are also more accurate.

In order to avoid confusion, the names of the methods or combination of methods used for construction the harmonic and anharmonic components of the PESs are summarize in Table 2.

Table 2. Nomenclature for the methods used to generate the PESs. Columns differ according to the method used to obtain the harmonic component force constants, while rows differentiate between electronic structure models used to compute anharmonic force constants. "Pure" force fields are those for which all force constants are computed at the same level of theory while "hybrid" force fields have their harmonic force constants literally replaced with reference values obtained at a higher level of theory.

	Harmonic force constants	
Force field method	Same as anharm.	Ref. Literature
HF	HF	HF-REF
MP2	MP2	MP2-REF
CCSD(T)	CCSD(T)	CCSD(T)-REF

3.3 Results and Discussion

3.3.1. Statistical analysis of anharmonic frequencies from approximate force field models

Errors were calculated as the absolute difference between the computed frequencies and the associated frequencies from the PyPES library (VCI(6)): $\Delta_v = \left|v_{ref} - v_{comp}\right|$. Mean, median, maximum and minimum absolute errors, together with the upper (Q3) and lower (Q1) quartiles and the interquartile range (IQR) in predicted fundamental frequencies according to approximate force field models are presented in Table 3 and illustrated in Figures 1 & 2. Note that all fundamental frequencies obtained using all pure and composite force fields for all molecules can be found in Appendix B.

Table 3. Errors between the computed fundamental frequencies and the associated reference frequencies, as well as the upper quartile, lower quartile and interquartile range for each method. All values are expressed in cm^{-1}.

	HF	MP2	CCSD(T)	HF - REF	MP2 - REF	CCSD(T) -REF
MEAN	155.3	49.7	11.0	21.7	8.6	2.6
MEDIAN	130.1	31.7	8.5	10.2	3.4	1.8
Q1	98.1	19.4	3.4	3.4	0.9	0.8
Q3	195.4	63.9	16.6	30.1	6.4	4.1
IQR	97.3	44.5	13.2	26.7	5.5	3.3
MAX	337.8	127.1	31.1	69.6	12.7	8.9
MIN	30.7	3.6	0.3	0.2	0.002	0.02

Figure 1. Magnitude and range of absolute errors in fundamental frequencies for each method are illustrated using the boxplot format. The median and lower/upper quartile are represented with the three horizontal lines that form each of the boxes. Maximum and minimum values are represented with the horizontal lines above and below every box and the individual crosses show the outliers of each method.

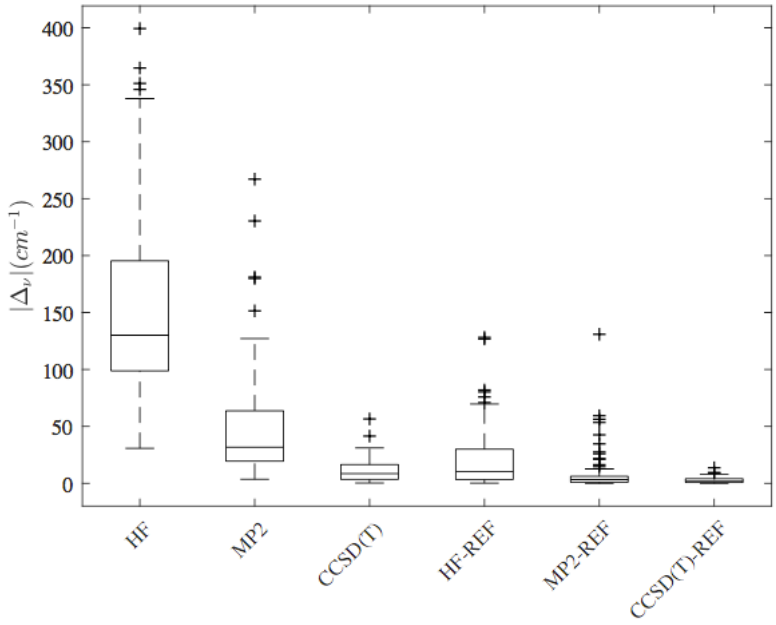
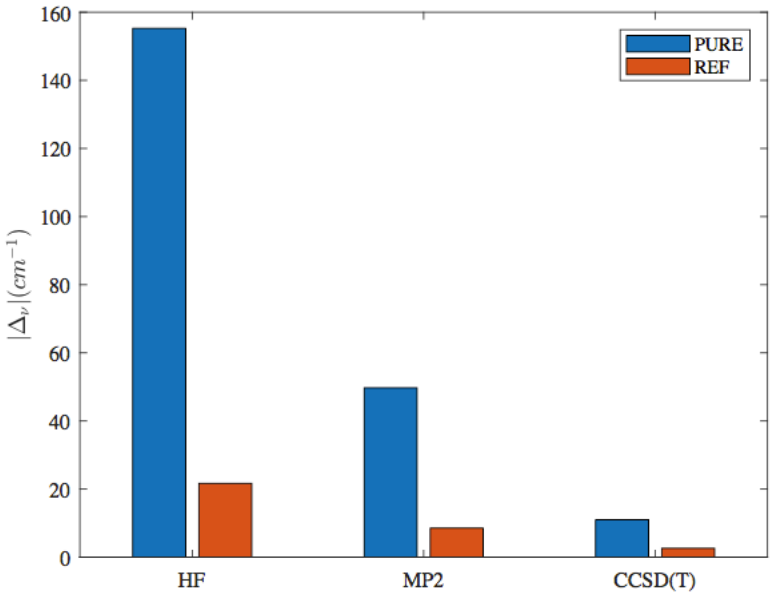


Figure 2. Average absolute errors in fundamental frequencies for each method. In blue, pure quartic force field and, in orange, harmonically replaced quartic force field. The harmonic force constants used for replacement were computed at, at least, CCSD(T)/cc-pVTZ level.



3.3.2. Statistical analysis of results from non-composite force fields

In the first part of this study, we constructed the PES using only the aforementioned electronic structure methods – HF, MP2 and CCSD(T) in conjunction with the cc-pVTZ atomic orbital basis set – and computed the fundamental frequencies for all the molecules using the VCI method. As expected, the most accurate results were obtained when using CCSD(T), with an absolute mean error of 11.0 cm^{-1} across all fundamental frequencies. From the dispersion in the errors, expressed as the interquartile range, we can infer that in 50% of the cases when assigning the experimental spectra using this method, the predicted values will fall within a range of 13.2 cm^{-1} from the median value. However, errors can increase more than 30 cm^{-1} in some cases.

Fundamental frequencies calculated using HF and MP2 force fields resulted in larger average errors and a wider distribution of errors, with an overall accuracy of 155.3 cm^{-1} and 49.7 cm^{-1} respectively. These large errors are expected because both HF and MP2 are substantially lower levels of theory than CCSD(T). There is approximately an order of magnitude difference in accuracy between the most accurate model – CCSD(T) – and the least accurate – HF. MP2 lies approximately half an order of magnitude in between.

3.3.3. Statistical analysis of results from composite force fields

In the second part of this investigation, harmonic force constants extracted from the reference force fields implemented within PyPES were introduced to the force fields generated using the methods stated above. The accuracy of the results improved significantly in all cases. With an absolute mean error of 21.7 cm^{-1}, the results obtained with the composite force field HF-REF are already comparable to those using solely CCSD(T) for the construction of the

entire force field. MP2-REF is already more accurate than CCSD(T) alone, with an overall average error below 10 cm^{-1}. Finally, by replacing the harmonic force constants in CCSD(T) with those from the reference literature, the errors decreased further, to 2.6 cm^{-1} for the absolute mean error.

As the average error decreases moving from HF-REF, through MP2-REF to CCSD(T)-REF so too does the spread of error values, assessed either in terms of the interquartile range or by looking at the maximum error. Overall, this confirms the reliability of the hybrid force field approach. However, some outliers can still be seen in Figure 1, which will be analysed in the following section.

3.3.4. Analysis of sources of error

The key tenet of this work is that the error distributions in Figure 1 reflect the quality of the underlying PES. Clearly, if the PES is generated using an accurate electronic structure model, i.e. CCSD(T)/cc-pVTZ or higher, then computed fundamentals closely match their reference values. However, by comparing CCSD(T) and CCSD(T)-REF results, it is also clear that the quality of the predictions is strongly linked to the accuracy of the harmonic part of the potential – average errors decrease from 11.0 cm^{-1} to 2.6 cm^{-1}, and maximum errors from 31.1 cm^{-1} to 8.9 cm^{-1}.

On the other hand, errors in composite force fields are expected to reflect the accuracy of the anharmonic force constants. However, for HF-REF and MP2-REF some extreme outliers occur at $\Delta_v > 100$ cm^{-1} that do not seem to fit in the same general distribution as the other error values. Fortunately, we can investigate these cases in more detail by inspecting the VCI coefficients of the corresponding vibrational modes. For completeness, we will also analyse other, less

extreme outliers with $\Delta_v > 70$ cm^{-1} (HF-REF), $\Delta_v > 15$ cm^{-1} (MP2-REF) and $\Delta_v > 9$ cm^{-1} (CCSD(T)-REF).

In some cases, the presence of outliers and larger errors is not related directly to the quality of the anharmonic force constants, but can arise indirectly due to Fermi resonances appearing in the wavefunction.[33] One such example is the third vibrational mode of C_4. The fundamental frequencies computed with the three composite force field methods used in this study, along with the VCI coefficients and the basis states contributors for each coefficient are presented in Table 5. We have also added the fundamental frequency extracted from the reference literature for comparison. Only the coefficients from the two VCI basis states that make the largest contribution to the overall wavefunction are included in the table.

Table 5. Frequency (cm^{-1}), VCI coefficients and basis state contributors for the third mode of C_4 computed using composite force fields (HF-REF, MP2-REF, CCSD(T)-REF). The first number in the state designation indicates the vibrational mode and the second represents its vibrational quantum number. The reference fundamental frequency was added for comparison.

	HF-REF		**MP2-REF**		**CCSD(T)-REF**		**Ref.**[13]
Freq.	923.0		972.6		930.8		931.9
VCI Coeff.	**Coef.**	**State/s**	**Coef.**	**State/s**	**Coef.**	**State/s**	
	0.928	(3,1)	0.609	(3,1)	0.911	(3,1)	
	0.043	(2,2)	0.328	(2,2)	0.046	(2,2)	

As we can see, the error obtained at MP2-REF is caused by an exaggerated resonance between two vibrational states: the third mode and an overtone of the second mode.

It is also possible to extract information about the nature of the outliers caused by the inaccuracy of the anharmonic force constants with the analysis of the VCI coefficients. Cases with a strong contribution from overtones of the same mode in the VCI wavefunction are due to inaccuracy in the diagonal force constants. HO_2^+, an outlier present in HF-REF and MP2-REF, and with an error above average in CCSD(T)-REF, is a good example of this and. Like in the previous case, we are providing the VCI coefficient data in Table 6.

Table 6. Frequency (cm^{-1}), VCI coefficients and basis state contributors for the third mode of HO_2^+ computed using composite force fields (HF-REF, MP2-REF, CCSD(T)-REF). The first number in the state designation indicates the vibrational mode and the second represents its vibrational quantum number. The reference fundamental frequency was added for comparison.

	HF-REF		**MP2-REF**		**CCSD(T)-REF**		**Ref.[13]**
Freq.	2902.0		3050.4		3022.9		3029.2
VCI Coeff.	**Coef.**	**State/s**	**Coef.**	**State/s**	**Coef.**	**State/s**	
	0.769	(3,1)	0.831	(3,1)	0.845	(3,1)	
	0.131	(3,2)	0.081	(3,2)	0.096	(3,2)	

In this instance, the wavefunction for the third vibrational mode includes an unexpectedly strong contribution from its own overtone, which causes a large error in the fundamental. This directly reflects the fact that the Hartree-Fock model must have overestimated the anharmonicity along this mode, i.e. predicted anharmonic force constants that are too large.

Errors arising from the off-diagonal force constants, which account for the coupling between modes, can be also distinguished by looking at the VCI wavefunction coefficients for a particular mode. Anharmonic force constants can be both under and over-estimated. If a particular force constant is underestimated we can expect a second contribution that is too weak, coming from a state that is formed by the same mode mixed with another mode. On the contrary, if the force constant is overestimated, the contribution will be too strong. The bending mode in H_2O computed at MP2-REF level is an example of the latter, as shown by the VCI coefficient data reported in Table 7.

Table 7. Frequency (cm^{-1}), VCI coefficients and basis state contributors for the first mode of H_2O computed using composite force fields (HF-REF, MP2-REF, CCSD(T)-REF). The first number in the state designation indicates the vibrational mode and the second represents its vibrational quantum number. The reference fundamental frequency was added for comparison.

	HF-REF		**MP2-REF**		**CCSD(T)-REF**		**Ref.[13]**
Freq.	1578.1		1613.2		1601.3		1596.8
VCI Coeff.	**Coef.**	**State/s**	**Coef.**	**State/s**	**Coef.**	**State/s**	
	0.974	(1,1)	0.947	(1,1)	0.980	(1,1)	
	0.016	(1,1)(2,1)	0.023	(1,1)(3,1)	0.014	(1,1)(2,1)	

Here, the first vibrational mode for MP2-REF has a slightly larger contribution from the mix of itself and the third mode, which causes an error in the fundamental frequency that makes it an outlier. In this case, the MP2 model must have overestimated the anharmonicity between the first and the third mode, i.e. predicted off-diagonal anharmonic force constants that are too large.

Lastly, we have cases whose wavefunctions do not present any unusual behaviour but the computed frequencies are inaccurate, which reflects more global and generalised errors in anharmonic force constants. NH_2^-, due to its anionic nature, is an example where this occurs. The VCI coefficients presented in Table 8 show no particular anomalous behaviour, but the computed fundamental frequencies are inaccurate across all three hybrid force field models. It is well known that diffuse functions are required in the basis set to accurately describe the electron distribution of anions. In this study, we have only used a cc-pVTZ basis set to compute the anharmonic force constants.

Table 8. Frequency (cm^{-1}), VCI coefficients and basis state contributors for the first mode of NH_2^- computed using composite force fields (HF-REF, MP2-REF, CCSD(T)-REF). The first number in the state designation indicates the vibrational mode and the second represents its vibrational quantum number. The reference fundamental frequency was added for comparison.

	HF-REF		**MP2-REF**		**CCSD(T)-REF**		**Ref.[13]**
Freq.	1436.7		1462.9		1461.8		1448.2
VCI Coeff.	**Coef.**	**State/s**	**Coef.**	**State/s**	**Coef.**	**State/s**	
	0.971	(1,1)	0.979	(1,1)	0.977	(1,1)	
	0.019	(1,1)(2,1)	0.015	(1,1)(2,1)	0.017	(1,1)(2,1)	

We have categorized all the outliers present in the cases where the composite force field approach was used, HF-REF, MP2-REF and CCSD(T)-REF, in Tables 9, 10 & 11, respectively. The classification is fourfold: inaccuracy of the anharmonic force constants (Anharmonic FCs, Diagonal FCs & Off-diagonal FC) or Fermi resonance

Table 9. Classification of the outliers present with (a) HF-REF force field, (b) MP2-REF force field and (c) CCSD(T)-REF force field. Fermi resonance errors occur when inaccurate anharmonic force constants introduce artificial resonances into the wavefunction that are not present at higher levels of theory. Errors in diagonal force constants (diagonal FCs) cause along-mode anharmonicity to be under or over-estimated and show up in the wavefunction through unusual patterns of overtone mixing. Errors in off-diagonal force constants (off-diagonal FCs) cause inaccurate mode-coupling which shows up in the wavefunction as unexpected contributions from combination bands. Global errors in anharmonic force constants do not show up clearly in the wavefunction but only in the values of the computed fundamentals.

(a) HF-REF

Molecule	Fundamental	Error	Explanation
C_4	1446.9	128.2	Fermi resonance
H_2O	3690.9	70.9	Off-diagonal FCs
HO_2^+	2902.4	126.8	Diagonal FCs
HOF	3528.9	75.8	Diagonal FCs
N_2H_2	2960.4	80.2	Fermi resonance
SiH_4	994.5	81.9	Off-diagonal FCs

(b) MP2-REF

Molecule	Fundamental	Error	Explanation
C_4	488.4	34.6	Off-diagonal FCs
	974.5	42.6	Fermi resonance
	1292.5	25.9	Fermi resonance
H_2O	1613.2	16.3	Off-diagonal FCs
	3602.4	59.4	Off-diagonal FCs
HO_2^+	1510.9	130.9	Off-diagonal FCs
	3050.4	21.2	Diagonal FCs
NH_2^-	1462.9	14.7	Anharmonic FC
SiH_4	934.2	21.7	Off-diagonal FCs
	996.6	27.5	Off-diagonal FCs
	2236.2	56.3	Diagonal FCs
	2244.1	53.5	Off-diagonal FCs

(c) CCSD(T)-REF

Molecule	Fundamental	Error	Explanation
C_3H_2	768.3	9.5	Off-diagonal FCs
NH_2^-	1461.8	13.6	Anharmonic FC

Fermi resonances only account for a very small proportion of the marked outliers in Figure 1, and the most extreme outliers are not due to Fermi resonances but instead genuinely reflect the quality of the underlying force field.

3.4. Conclusions

Every investigation in the field of computational chemistry is subject to the computing power available. In cases where the computational resources are abundant and high accuracy is desired, we recommend using CCSD(T) with the largest possible basis set, at least aug-cc-pVTZ, for constructing the entire force field expansion. Unfortunately, this is not often the case. The commonly used composite force field approach can be a computationally efficient alternative to obtain high quality results when predicting the fundamental frequencies. We hope the results of this study can be interpreted by computational chemists as a reference of the errors that can be expected when using this approach.

We have seen how the largest contribution of error in the calculation of those frequencies stems from the harmonic part of the PES. Therefore, the harmonic force constants should be generated at the highest possible level of theory and largest possible basis set. After that, it becomes a matter of matching the accuracy of the harmonic part against the accuracy of the anharmonic part of the force field in order to obtain the desired accuracy in the results.

Within the composite force field framework, we recommend, in cases where more accuracy is desired, using CCSD(T) together with, at least, a cc-pVTZ basis set for the construction of the harmonic part of the expansion of the force field and MP2 with cc-pVTZ for the anharmonic counterpart. This approach incur an absolute error in the fundamental frequencies of approximately 10 cm^{-1} when compared against "spectroscopically accurate" computational values reference that themselves are within 5 cm^{-1} of experiment. With this composite force field selection, accuracy is not compromised and computing cost is reduced considerably when compared to full CCSD(T).

If the accuracy of the results is not a matter of paramount importance, constructing the force field expansion at HF level and replacing the harmonic force constants by those computed at CCSD(T)/cc-pVTZ (or above), could represent a good alternative for doing semi-quantitative and qualitative assignments of vibrational spectra, yielding errors of 20 cm^{-1} approximately.

The presence of outliers in the results using the composite force field approach arose directly from intrinsic inaccuracy of the electronic structure model used in computing anharmonic force constants. In some cases, artefactual Fermi resonances can be a problem, but these too indirectly reflect the quality of the underlying PES.

References

(1) Christiansen, O. Vibrational structure theory: New vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties. *Physical Chemistry Chemical Physics* **2007**, *9*, 2942-2953.

(2) Espinosa-Garcia, J.; Monge-Palacios, M.; Corchado, J. C. Constructing Potential Energy Surfaces for Polyatomic Systems: Recent Progress and New Problems. *Advances in Physical Chemistry* **2012**, *2012*, 164752.

(3) Krasnoshchekov, S. V.; Craig, N. C.; Koroleva, L. A.; Stepanov, N. F. Anharmonic vibrational analysis of s-trans and s-cis conformers of acryloyl fluoride using numerical-analytic Van Vleck operator perturbation theory. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2018**, *189*, 66-79.

(4) Pietropolli Charmet, A.; Bizzocchi, L.; Giuliano, B. M.; Caselli, P.; Craig, N. C.; Krasnoshchekov, S. V. Disentangling the IR spectra of 2,3,3,3-tetrafluoropropene using an ab initio description of vibrational polyads by means of canonical Van Vleck perturbation theory. *Journal of Quantitative Spectroscopy and Radiative Transfer* **2019**, *239*, 106656.

(5) Puzzarini, C.; Biczysko, M.; Barone, V. Accurate Harmonic/Anharmonic Vibrational Frequencies for Open-Shell Systems: Performances of the B3LYP/N07D Model for Semirigid Free Radicals Benchmarked by CCSD(T) Computations. *Journal of Chemical Theory and Computation* **2010**, *6*, 828-838.

(6) Begue, D.; Carbonniere, P.; Pouchan, C. Calculations of Vibrational Energy Levels by Using a Hybrid ab Initio and DFT Quartic Force Field: Application to Acetonitrile. *The Journal of Physical Chemistry A* **2005**, *109*, 4611-4616.

(7) Puzzarini, C.; Biczysko, M.; Barone, V. Accurate Anharmonic Vibrational Frequencies for Uracil: The Performance of Composite Schemes and Hybrid CC/DFT Model. *Journal of Chemical Theory and Computation* **2011**, *7*, 3702-3710.

(8) Ramachandran, K.; Deepa, G.; Namboori, K.: *Computational chemistry and molecular modeling: principles and applications*; Springer Science & Business Media, 2008.

(9) Davisson, J. L.; Brinkmann, N. R.; Polik, W. F. Accurate and efficient calculation of excited vibrational states from quartic potential energy surfaces. *Molecular Physics* **2012**, *110*, 2587-2598.

(10) Wilson, E. B.: Molecular vibrations: the theory of infrared and Raman vibrational spectra; *McGraw-Hill* **1955**.

(11) Bowman, J. M.; Bowman, J. M. The self-consistent-field approach to polyatomic vibrations. *Accounts of Chemical Research* **1986**, *19*, 202-208.

(12) Sibaev, M.; Crittenden, D. L. PyVCI: A flexible open-source code for calculating accurate molecular infrared spectra. *Computer Physics Communications* **2016**, *203*, 290-297.

(13) Sibaev, M.; Crittenden, D. L. The PyPES library of high quality semi-global potential energy surfaces. *Journal of computational chemistry* **2015**, *36*, 2200-2207.

(14) Sibaev, M.; Crittenden, D. L. An efficient and numerically stable procedure for generating sextic force fields in normal mode coordinates. *The Journal of Chemical Physics* **2016**, *144*, 214107.

(15) Delahaye, T.; Nikitin, A.; Rey, M.; Szalay, P. G.; Tyuterev, V. G. A new accurate ground-state potential energy surface of ethylene and predictions for rotational and vibrational energy levels. *Journal of Chemical Physics* **2014**, *141*, 104301.

(16) Dateo, C. E.; Lee, T. J. An accurate ab initio quartic force field and vibrational frequencies for cyclopropenylidene. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53*, 1065-1077.

(17) Wang, X.; Huang, X.; Bowman, J. M.; Lee, T. J. Anharmonic rovibrational calculations of singlet cyclic C4 using a new ab initio potential and a quartic force field. *The Journal of chemical physics* **2013**, *139*, 224302.

(18) Lee, T. J.; Martin, J. M. L.; Taylor, P. R. An Accurate ab initio Quartic Force Field and Vibrational Frequencies for CH4 and Isotopomers. *Journal of Chemical Physics* **1995**, *102*, 254-261.

(19) Breidung, J.; Thiel, W.; Gauss, J.; Stanton, J. F. Anharmonic force fields from analytic CCSD(T) second derivatives: HOF and F2O. *The Journal of Chemical Physics* **1999**, *110*, 3687-3696.

(20) Yachmenev, A.; Jensen, P.; Yurchenko, S.; Thiel, W. A new "spectroscopic" potential energy surface for formaldehyde in its ground electronic state. *Journal of Chemical Physics* **2011**, *134*, 244307-244307-244311.

(21) Huang, X.; Lee, T. J. A procedure for computing accurate ab initio quartic force fields: Application to HO2+ and H2O. *The Journal of chemical physics* **2008**, *129*, 044312.

(22) Peterson, K. A. Accurate ab initio near-equilibrium potential energy and dipole moment functions of HOCl and HOBr. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53*, 1051-1064.

(23) Martin, J. M. L.; Taylor, P. R. Accurate ab initio quartic force field for trans-HNNH and treatment of resonance polyads. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53*, 1039-1050.

(24) Craig, N. C.; Appiah, K. J.; Miller, C. E.; Seiden, M. V.; Varley, J. E. Reevaluation of matrix-isolation infrared spectra of the isotopologues of trans-diazene and attempts to prepare cis-diazene by photoisomerization. *Journal of Molecular Spectroscopy* **2015**, *310*, 3-7.

(25) Huang, X.; Lee, T. J. Accurate ab initio quartic force fields for NH[sub 2]−] and CCH[sup −] and rovibrational spectroscopic constants for their isotopologs. *The Journal of Chemical Physics* **2009**, *131*, 104301.

(26) Yurchenko, S. N.; Zheng, J.; Lin, H.; Jensen, P.; Thiel, W. Potential-energy surface for the electronic ground state of NH3 up to 20,000 cm-1 above equilibrium. *The Journal of chemical physics* **2005**, *123*, 134308.

(27) Martin, J. M. L.; Lee, T. J. Accurate ab initio quartic force field and vibrational frequencies of the NH4+ ion and its deuterated forms. *Chemical Physics Letters* **1996**, *258*, 129-135.

(28) Ovsyannikov, R. I.; Thiel, W.; Yurchenko, S. N.; Carvajal, M.; Jensen, P. Vibrational energies of PH3 calculated variationally at the complete basis set limit. *The Journal of Chemical Physics* **2008**, *129*, 044309.

(29) Pak, Y.; Woods, R. C. Coupled cluster calculations of the potential energy surfaces and spectroscopic constants of SiF2, PF+2, SO2, PO−2, and ClO+2. *The Journal of Chemical Physics* **1996**, *104*, 5547-5554.

(30) Martin, J. M. L.; Baldridge, K. K.; Lee, T. J. Accurate ab initio anharmonic force field and heat of formation for silane. *Molecular Physics* **1999**, *97*, 945.

(31) Martin, J. L. A fully ab initio quartic force field of spectroscopic quality for SO3. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1999**, *55*, 709-718.

(32) J.F. Stanton, G., J., Cheng, J., Harding, M.E., Matthews, D.A., Szalay, P.G.,; with contributions from Auer, A. A., Bartlett, R.J., Benedikt, U., Berger, C.,; Bernholdt, D. E., Bomble, Y.J.,Christiansen, O., Engel, F., Faber, R., Heckert,; M., H., O., Hilgenberg, M., Huber, C., Jagau, T.-C., Jonsson, D., Jusélius, J.,; Kirsch, T., Klein, K., Lauderdale, W.J., Lipparini, F., Metzroth T., Mück, L.A., ; 21; O'Neill, D. P., Price, D.R., Prochnow, E., Puzzarini, C., Ruud, K.,Schiffmann, F.,; Schwalbach, W., Simmons, C., Stopkowicz, S., Tajti, A., Vázquez, J., Wang, F.,; Watts, J. D. a. t. i. p. M. A., J. and Taylor,; P.R.), P. T., P.R.), ABACUS (Helgaker, T., Jensen, H.J.Aa.,; Jørgensen, H. P., and Olsen, J.), and ECP routines by Mitin, A.V. and Wüllen C.; van, C., Coupled-Cluster techniques for Computational Chemistry, a; (2018)., q.-c. p. p.: CFOUR , Coupled-Cluster techniques for Computational CHemistry. 2018.

(33) Schrader, B.: *Infrared and Raman spectroscopy: methods and applications*; John Wiley & Sons, 2008.

Chapter 4

A Discrete Dichloride Tetrahydrate Trapped by a Cyclopropenium Cation: Structure and Spectroscopic Properties

---

"Observe constantly that all things take place by change, and accustom thyself to consider that the nature of the Universe loves nothing so much as to change things which are, and to make new things like."

Marcus Aurelius

The material contained within this chapter has previously been published in a peer-reviewed journal article. The citation for this article is stated below.

- Abdelbassit, M. S.; Curnow, O. J.; Ferreras, M.; Crittenden, D. L. A Discrete Dichloride Tetrahydrate Trapped by a Cyclopropenium Cation: Structure and Spectroscopic Properties. *ChemPlusChem* **2020**, *85*, 927-932.

The contributions of each author are the following. Mohammed Abdelbassit was responsible for characterising and synthesising the materials as well as writing the manuscript together with Deborah Crittenden and Owen Curnow. The computational design was created by Deborah Crittenden and Manuel Ferreras Moreno, who also conducted all the simulations. The experimental design was directed by Owen Curnow.

Dr. Matthew I. J. Polson (University of Canterbury) is thanked for his assistance with the X-ray crystallography.

4.1. Introduction

Chloride-water clusters are fundamentally-important motifs for understanding a wide range of physicochemical processes in nature. For example, they play a vital role in biological and geological systems; particularly at water/membrane interfaces and in the mobility of ions through ion channels, as well as in electrical phenomena in the troposphere and ionosphere.[1-3] Indeed, there are extensive theoretical studies on monochloride hydrate $[Cl(H_2O)_n]^-$ clusters over the last several years.[4-18] On the other hand, there are only five reported *ab initio* studies on dichloride hydrates $[Cl_2(H_2O)_n]^{2-}$.[19-25] Despite the expected ion-ion repulsion between chloride ions, there is strong evidence for the formation of Cl^----Cl^- ion pairs in aqueous solution, particularly in concentrated solutions.[19-21,26-31]

In 2013, Basu et al. synthesized the simplest hydrated dichloride cluster found to date, $[Cl_2(H_2O)_2]^{2-}$, by using an amide receptor. The Cl^- anions are bridged by two H_2O molecules, and the structure is stabilized by additional hydrogen bonds (two NH---OH_2 and two NH---Cl^-).[32] $[Cl_2(H_2O)_2]^{2-}$ was also reported by Szumna et al. in 2001 as a co-crystal of a neutral octalactam receptor with tetrabutylammonium chloride. The $[Cl_2(H_2O)_2]^{2-}$ cluster lies in a monocyclic cavity held via eight hydrogen bonds (four NH---OH_2 and four NH---Cl^-). There is also hydrogen bonding between the Cl^- and dichloromethane in the extended lattice.[33] In 2012, Wang et al. prepared a discrete $[Cl_2(H_2O)_2]^{2-}$structure; although it exhibits strong Cl^-–π and H_2O–π interactions with electron-deficient regions of the triazine rings.[34] Furthermore, Safin et al. claimed a discrete structure of dichloride dihydrate, however, the cluster is in fact supported by four amino-NH hydrogen bonds with H_2O and Cl^-.[35] We do not believe that a cluster held together by hydrogen bonding should be described as "discrete" when ignoring additional strong hydrogen bonds to the cluster fragment.

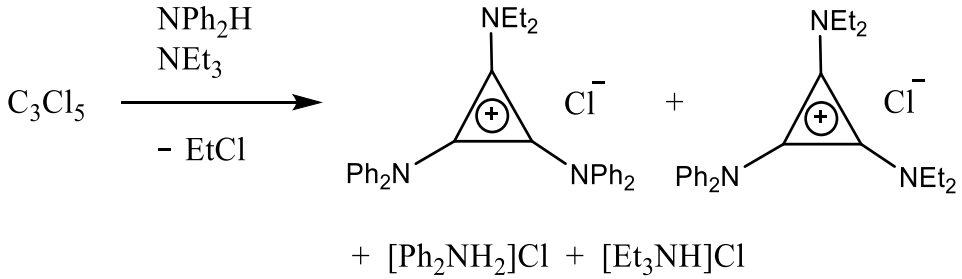
To the best of our knowledge there are only three solid-state structures of the tetrahydrate $[Cl_2(H_2O)_4]^{2-}$ that have been reported. The first one has prepared as the counterion of a nickel(II) complex and is attached to the ligands through six NH···Cl^- and four NH···OH_2 hydrogen bonds.[36] The second structure, reported by Arunachalam and Ghosh, was formed in two bowl-shaped tripodal receptors in an octahedral-like arrangement in which all of the water molecules are disordered over two positions and which consequently defies a reasonable structural description. The structure is stabilized by multiple amide –NH hydrogen bonding with Cl^- and H_2O.[37] The third structure was reported by Chakraborty et al. in 2014 as a tetrabutylammonium salt; the hydrated cluster is encapsulated in a pyridyl-functionalized receptor. Nevertheless, the structure shows strong hydrogen bonding between the cluster and the receptor.[38] We have previously reported the structure of a discrete dichloride hexahydrate $[Cl_2(H_2O)_6]^{2-}$ cube with the triaminocyclopropenium (TAC) cation $[C_3(N^iPr_2)_3]^+$.[23,24] The larger dichloride-water aggregates $[Cl_2(H_2O)_8]^{2-}$[39] and $[Cl_2(H_2O)_{14}]^{2-}$[40] have been investigated in supramolecular systems as part of hydrogen-bonding networks. The decahydrate $[Cl_2(H_2O)_{10}]^{2-}$ has been obtained as one unit, but with strong hydrogen bonding to two adjacent Na^+-water clusters.[41] Unfortunately, the vibrational modes were not reported for any of these clusters due to overlapping with the bands of the NH groups, except for the dichloride hexahydrate.[23,24] The vibrational modes of a limited number of monochloride hydrates have been observed via Ar-tagging at low temperature in the gas phase.[17,18]

Herein, we present a new hydrophobic environment, $[C_3(NPh_2)_2(NEt_2)]^+$, to host a discrete dichloride tetrahydrate cluster. The crystal structure and the spectroscopic properties (FT-IR) of $[Cl_2(H_2O)_4]^{2-}$ and the $[Cl_2(D_2O)_4]^{2-}$ analogue, as well as a theoretical study, will be discussed.

4.2. Methods

4.2.1. Synthesis

We synthesized the new TAC cation diethylamino-bis(diphenylamino)cyclopropenium, $[C_3(NPh_2)_2(NEt_2)]^+$, as a chloride salt via the addition of a mixture of $HNPh_2$ and NEt_3 (3 : 4.5 molar ratio) to one equivalent of pentachlorocyclopropane (Scheme 1). Four cations were obtained: the ammonium salts $[H_2NPh_2]Cl$ and $[HNEt_3]Cl$ along with two TAC-based salts $[C_3(NEt_2)_2(NPh_2)]Cl$ and $[C_3(NPh_2)_2(NEt_2)]Cl$, from which the latter was ultimately isolated as a dihydrate in 41% yield. The crystallization was carried out in a water-saturated binary solvent of dichloromethane and ethyl acetate (1:1 v/v) at room temperature. The $[Cl_2(D_2O)_4]^{2-}$ analogue was prepared by evaporation of a D_2O-saturated dichloromethane solution.



Scheme 1. Synthesis of $[C_3(NPh_2)_2(NEt_2)]Cl$.

4.2.2. Computational

Geometry optimizations and harmonic vibrational frequency calculations were performed on gas-phase $[Cl_2(H_2O)_4]^{2-}$ clusters at HF/pc-2, HFS/pc-2, B3LYP/6-31+G*, MP2/aug-cc-pVDZ and using a custom density functional model comprising 50% HF

exchange and 50% Slater exchange with a pc-2 basis set (HnH/pc-2). An empirical quadratic correction model was applied to obtain predicted anharmonic vibrational frequencies.[42]

4.3. Results and Discussion

4.3.1. Single crystal structure

$[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$ crystallizes in the triclinic space group *P-1* (Table 1) in which the asymmetric unit consists of one cation and a chloride dihydrate. In the cation, the phenyl groups exhibit a propeller conformation to minimize the steric interactions (Figure 1).

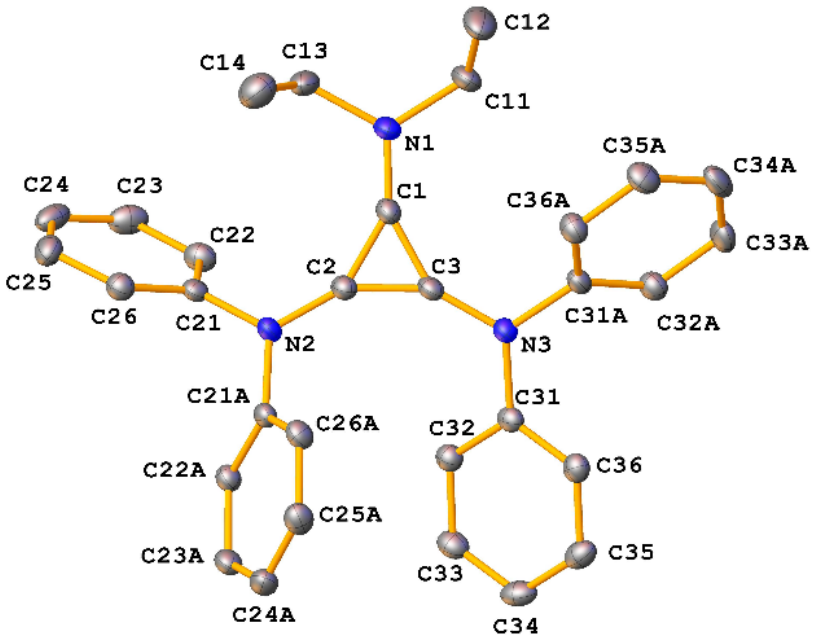


Figure 1. The geometry and labeling scheme of the cation in $[C_3(NPh_2)_2(NEt_2)]Cl{\cdot}2H_2O$. Hydrogen atoms are omitted for clarity

Table 1. Structure refinement data for $[C_3(NPh_2)_2(NEt_2)]Cl.(H_2O)_2$.

formula	$C_{31}H_{34}ClN_3O_2$
a [Å]	8.9507(3)
b [Å]	11.4993(3)
c [Å]	14.6291(4)
α [°]	68.871(3)
β [°]	79.571(2)
γ [°]	89.225(2)
V [Å^3]	1379.10(7)
Z	2
ρ_{calc} [g cm^{-3}]	1.243
crystal system	triclinic
space group	P-1
shape/colour	colourless
crystal size [mm]	0.299 × 0.243 × 0.113
μ [mm^{-1}]	1.476
T [K]	120(10)
F(000)	548
2θ range [°]	8.256 – 147.114
index ranges	$-11 \le h \le 11$
	$-13 \le k \le 14$
	$-18 \le l \le 18$
reflections collected	28951
independent reflns	5536
R(int)	0.0304
data/restraints/parameters	5536/2/346
GoF on F^2	1.063
R_1/wR_2 [$I>2\sigma(I)$]	0.0429/0.1181
R_1/wR_2 (all data)	0.0469/0.1215
$\Delta\rho_{max/min}$/e [Å^{-3}]	1.19/–0.47

The structural parameters for the cation are given in the Appendix C. Bond distances indicate that the π donation into the C_3 ring from the diethylamino group is enhanced by the presence of the electronegative phenyl groups. The extended lattice shows that the cations form a short contact through the phenyl groups (C36'---C36") of 3.135(3) Å (Figure 2), which is found to be shorter than most "π-stacked" neutral arenes at 3.3–3.8 Å.[43] This strong interaction can be attributed to the substituent effects on the π-stacking geometry and thus the formation of benzene dimers.[44-46]

The chloride dihydrate forms a C_i-symmetric dimeric structure of $[Cl_2(H_2O)_4]^{2-}$ which can be described as having a square $[Cl_2(H_2O)_2]^{2-}$ core with two additional terminal H_2O bridging molecules on opposite Cl---O edges. Folding about the bridged edges produces a chair-like conformation (Figure 3).

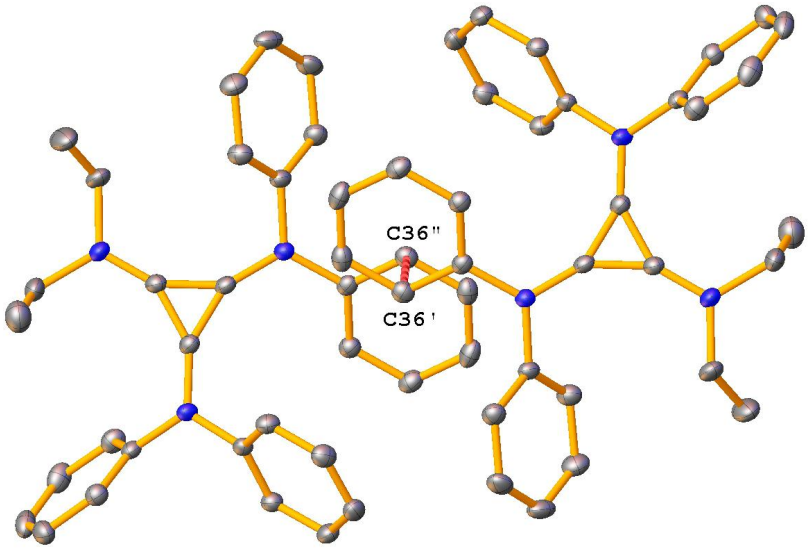


Figure 2. Edge-to-edge π-stacking between two cations in $[C_3(NPh_2)_2(NEt_2)]Cl{\cdot}2H_2O$, C36'---C36" 3.135(3) Å. Hydrogen atoms are omitted for clarity.

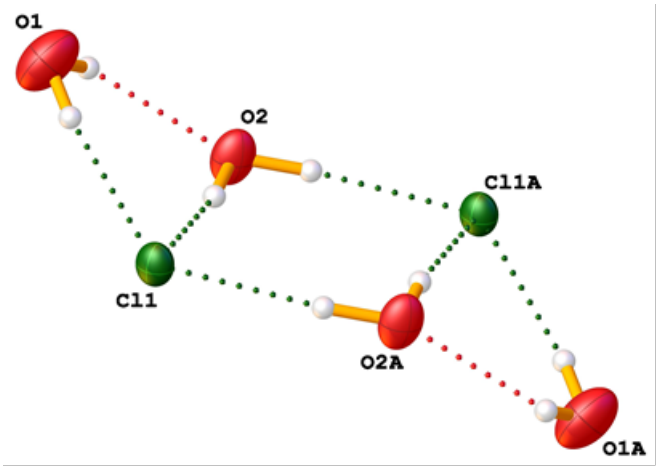


Figure 3. The structure of $[Cl_2(H_2O)_4]^{2-}$ in $[C_3(NPh_2)_2(NEt_2)]Cl{\cdot}2H_2O$. Dashed lines indicate hydrogen bonds.

Figure 4 depicts the environment around the $[Cl_2(H_2O)_4]^{2-}$ cluster. The cluster is sandwiched between two cyclopropenium cations and supported by another two cations via weak hydrogen bonds (C---O ~3.4 Å) between the water molecules within the cluster and neighbouring aryl and methyl substituents (see Figure 1S for detailed illustration). Notably, the chloride anions are located even further away from the centers of the cyclopropenium rings (Cl---C_3 = 3.48 Å). Each chloride anion engages in six hydrogen bonds; three of them are strong Cl---HO hydrogen bonds within the cluster (3.19 – 3.29 Å) and three are weak Cl---HC aryl hydrogen bonds to the surrounding cations (3.64 – 3.67 Å). This contrast between strong intra-cluster hydrogen bonding and weak interactions with the surrounding cyclopropenium units provides strong evidence for the existence of a discrete dichloride tetrahydrate structure within the crystalline environment.

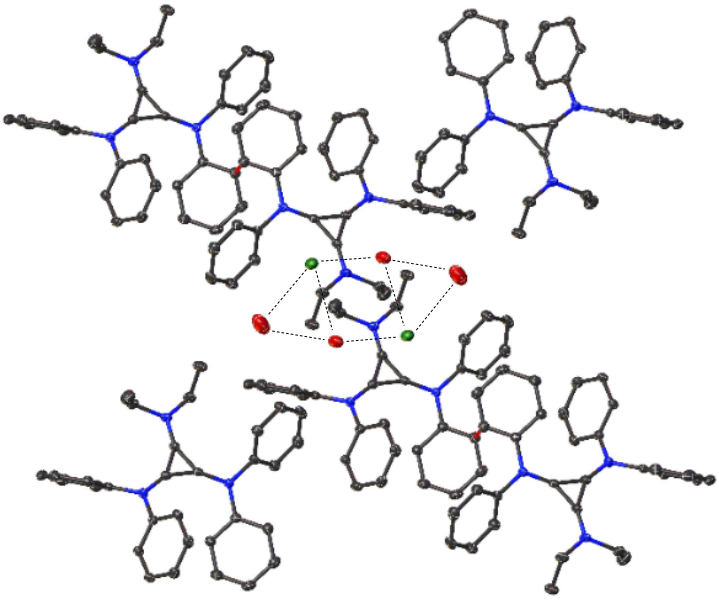


Figure 4. The environment of $[Cl_2(H_2O)_4]^{2-}$ in $[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$ Hydrogen atoms are omitted for clarity.

If the cluster we have identified is truly discrete, then it should be stable or at least semi-stable in the gas phase. To the best of our knowledge, no direct experimental evidence for gas phase

$[Cl_2(H_2O)_4]^{2-}$ clusters is available to date, but there is strong computational evidence that such clusters can exist in thermodynamically stable forms in the gas phase.[22,25] However, previous studies have either failed to identify the C_i-symmetric cluster as a local minimum,[25] or failed to disclose pertinent structural and spectroscopic data for this particular conformation.[22] Although previous studies indicate that the C_i-symmetric cluster is not the global minimum, it is reasonable to expect that the greater symmetry of the C_i cluster would be favoured during crystal formation over the C_1-symmetric extended bridged structures that would otherwise form to minimize ion-ion repulsion in the gas phase.[22] Here we report predicted structural parameters for the C_i-symmetric cluster in Table 2, along with conservative computational error estimates for the interatomic distance parameters. For reference, structural parameters for other crystallographically-characterized $[Cl_2(H_2O)_4]^{2-}$ and $[F_2(H_2O)_4]^{2-}$ non-discreet clusters have also been collated from the literature.[36,38]

Table 2. Structural parameters for $[X_2(H_2O)_4]^{2-}$ (X = F, Cl) clusters.

Parameter	X-ray	Calculated best estimate[a]	Calculated range [b]	HIMJUJ[36]	CIVGAR[38]	CIVGEV[38]
X	Cl	Cl	Cl	Cl	Cl	F
O1---O2 [Å]	3.116(3)	2.92	2.71 – 3.11	2.707(11)	4.456(3)	4.0067(7)
X–X [Å]	5.0932(8)	5.50	5.09 – 5.97	5.175(4)	4.317(13)	3.961(5)
X1---O1 [Å]	3.191(2)	3.29	3.17 – 3.44	4.015(6)	3.251(2)	2.647(7)
X1---O2 [Å]	3.2900(16)	3.34	3.14 – 3.62	3.172(6)	3.375(3)	2.702(5)
X1A---O2 [Å]	3.2250(17)	3.34	3.13 – 3.59	3.157(8)	3.317(3)	2.618(6)
X1–O2–X1A [°]	102.84(5)	110.9	--	109.7(2)	82.97(6)	97.21(17)
O2–X1–O2A [°]	77.16(5)	69.1	--	70.3(2)	97.03(6)	83.79(17)
O1–X1–O2 [°]	57.46(5)	52.2	--	42.24(18)	88.45(7)	99.1(2)
O2–O1–X1 [°]	62.86(5)	64.8	--	52.00(16)	44.72(5)	40.19(12)
O1–O2–X1 [°]	59.68(5)	63.0	--	85.8(2)	46.82(5)	40.73(14)
Fold angle [°]	29.38(7)	50.3	--	40.4(3)	17.41(6)	60.6(2)

[a] MP2/aug-cc-pVDZ, [b] LDA/pc-2 and HF/pc-2 values provide conservative lower and upper bounds for predicted interatomic distances.[47]

Direct comparison of computed and observed structural parameters is not generally straightforward, because differences between them can be attributed to a combination of physical effects (crystal packing, thermal and zero-point vibrational motion) and computational artefacts (errors associated with choice of electronic structure method and basis set incompleteness) that cannot be disentangled. Nonetheless, it is clear that only the $[Cl_2(H_2O)_4]^{2-}$ structure reported here resembles its gas phase counterpart, according to the parameters reported in Table 2. For all intra-cluster distances, our observed values fall within the computed upper and lower bounds, although they are not in perfect agreement with the best computational estimates. This implies that crystal packing and nuclear vibrational effects are of approximately the same magnitude as computational method uncertainties.

Comparison with other literature structures reveals that much larger differences are observed if the cluster forms additional bonds with surrounding molecules. Kopylovich et al. synthesized similar structures of a hexameric water cluster $(H_2O)_6$ and a dichloride tetrahydrate $[Cl_2(H_2O)_4]^{2-}$ (CSD reference code HIMJUJ), which contain a square core of $(H_2O)_4$ and $[Cl_2(H_2O)_2]^{2-}$, respectively, with two additional H_2O molecules.[36] Their structure is similar to the one reported here, and consistent with gas phase computational data except for the extended Cl---O1 distance of 4.02 Å, which arises from the fact that the terminal or "outer" water does not form a hydrogen bond back to chloride but instead hydrogen bonds to a neighbouring pyridine ring (N---O = 2.81 Å).

Chakraborty et al. reported the structures $[Cl_2(H_2O)_4]^{2-}$ (CIVGAR) and $[F_2(H_2O)_4]^{2-}$ (CIVGEV),[38] but these clusters are even less discrete than those reported by Kopylovich et al. Both are stabilized by strong (amide)-NH hydrogen bonding from the receptor to Cl^- and H_2O. Notably, the terminal water molecules hydrogen bond to both the pyridine ring in the receptor

and the chloride ions, which prevents them interacting with the central water molecules (O---O = 4.46 Å).

Further evidence that our crystallographically-trapped cluster resembles its gas phase counterpart lies in the close agreement between computed and observed intermolecular bonding angles, which agree to within 10° for our system but differ substantially in others.

At first glance, the ~20° difference between fold angles that the terminal water molecule subtends to the central dichloride dihydrate square appears to contradict the otherwise close agreement between computed gas phase structural parameters and their experimentally observed X-ray structure counterparts. However, this hinge angle is particularly soft and therefore can be easily influenced by weak crystal-packing effects.

4.3.2. Infrared Spectroscopy

Vibrational spectroscopy provides detailed information about bonding within and around solvated ion complexes,[47] and therefore is an ideal technique to probe the discreteness of the dichloride tetrahydrate captured here.

Mid-IR spectra were recorded for $[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$ and its deuterated analogue $[C_3(NPh_2)_2(NEt_2)]Cl.2D_2O$, and subtracted from one another to suppress bands arising from the cyclopropenium units and isolate peaks arising from the dichloride tetrahydrate clusters. The stretching region of the resultant difference spectrum is illustrated in Figure 5. The band patterns for the isotopomers are identical, but shifted due to isotopic mass effects. Peak positions are reported in Table 3. Similar analysis of the bending region reveals two bending bands – a strong band at 1625 cm^{-1} (1198 cm^{-1} for deuterated isotopologue) and a much weaker

band at 1655 cm^{-1} (1225 cm^{-1} for deuterated isotopologue). This immediately helps identify and assign the peaks at 3250 cm^{-1} and 3310 cm^{-1} as bending overtones.

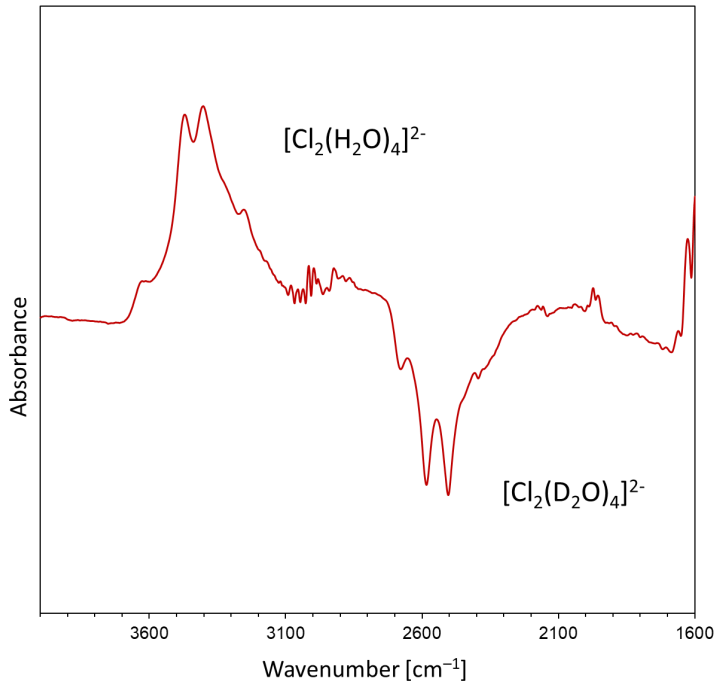


Figure 5. FT-IR spectrum of $[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$ minus the spectrum of $[C_3(NPh_2)_2(NEt_2)]Cl.2D_2O$.

To make further progress in assigning the observed spectrum, we turn to the computational results reported in Table 3 and illustrated in Figure 6. A pair of high frequency asymmetric stretching modes separated by ~ 20 cm^{-1} are expected, but only one is observed. It is assigned to the fundamental stretching mode that undergoes the largest change in dipole moment; ν(asym, syn). It is most likely that the nearby lower-intensity ν(asym, anti) fundamental is simply subsumed into the band centred at 3632 cm^{-1}.

Table 3. Experimental and calculated anharmonic vibrational frequencies (in cm^{-1}) for $[Cl_2(H_2O)_4]^{2-}$ and $[Cl_2(D_2O)_4]^{2-}$

Solid-state experimental	Gas phase calculated[a]	Deuterated experimental	Assignment
3632	3475–3594	2677	ν(asym, syn)
–	3451–3572	–	ν(asym, anti)
3470	3418–3528	2583	ν(sym, outer)
3405	3364–3490	2504	ν(sym, inner)
3310	3356–3472	2453	2ν(bend, outer)
3250	3192–3314	2394	2ν(bend, inner)
1655	1678–1736	1225	ν(bend, outer)
1625	1596–1657	1198	ν(bend, inner)

[a] Range of frequencies computed at B3LYP/6-31+G*, MP2/aug-cc-pVDZ and HnH/pc-2, with empirical anharmonic correction applied ($\Delta_{anh} = 0.011214\nu_h + 0.000010982\nu_h^2$)

The two symmetric stretching fundamentals at 3470 cm^{-1} and 3405 cm^{-1} are easy to identify and assign, because they are strong, high intensity transitions whose band centres fall right in the middle of their computationally-predicted ranges.

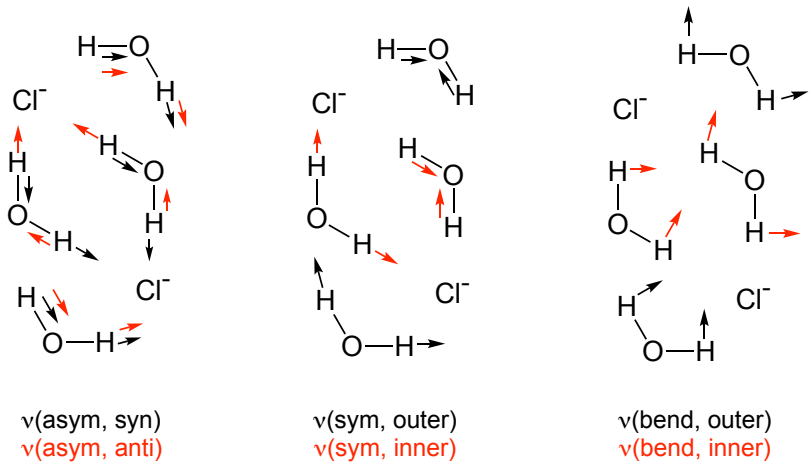


Figure 6. Fundamental stretching and bending vibrational modes of dichloride tetrahydrate.

The highest intensity bending mode corresponds to the in-phase bending of the two "inner" water molecules within the square planar dichloride dihydrate core. Both the fundamental and overtone bands are clearly visible in the difference spectrum at 1625 and 3250 cm^{-1}, respectively. Bending of the outer water molecules causes a much smaller change in dipole

moment of the dichloride tetrahydrate cluster as a whole, so the ν(bend, outer) fundamental has a much lower transition intensity and is observed as a very weak band at 1655 cm^{-1}. Its overtone is barely visible and only shows up as a very weak shoulder at 3310 cm^{-1}.

A few small discrepancies between theory and experimental exist; ν(bend, outer) appears to be slightly overestimated by all computational approaches employed here and ν(asym, syn) slightly underestimated. Unlike for the structural parameters, we have not attempted to provide strict bounds for the predicted anharmonic frequencies, so it is certainly plausible that these remaining discrepancies are purely computational artefacts. However, it is also possible that crystal packing and thermal effects that we have not attempted to account for here could account for the residual differences between computational predictions and experimental observations.

In either case, the differences are small, and inconsequential. The close agreement between computationally-predicted anharmonic gas phase transitions frequencies and observed solid state IR band centres provides strong evidence that we have isolated a discrete dichloride tetrahydrate structure, encapsulated predominantly through steric confinement, i.e. exhibiting only weak interactions with its surrounding environment.

4.4. Conclusions

We have illustrated the first example of a discrete dichloride tetrahydrate, $[Cl_2(H_2O)_4]^{2-}$, in the solid-state by taking advantage of the weak interactions between anions and triamino-cyclopropenium cations. The vibrational modes of a discrete chloride hydrate species have been observed and assigned for the first time in the solid state. In addition, we have also

reported the optimized structure and the calculated IR-active vibrational frequencies of $[Cl_2(H_2O)_4]^{2-}$ in the gas-phase which are in excellent agreement with the solid state structure and the experimental infrared spectrum.

References

(1) Castleman, A.; Bowen, K. Clusters: Structure, energetics, and dynamics of intermediate states of matter. *The Journal of Physical Chemistry* **1996**, *100*, 12911-12944.

(2) Ohtaki, H.; Radnai, T. Structure and dynamics of hydrated ions. *Chemical Reviews* **1993**, *93*, 1157-1204.

(3) Roux, B.; Bernèche, S.; Egwolf, B.; Lev, B.; Noskov, S. Y.; Rowley, C. N.; Yu, H. Ion selectivity in channels and transporters. *Journal of general physiology* **2011**, *137*, 415-426.

(4) Bajaj, P.; Zhuang, D.; Paesani, F. Specific Ion Effects on Hydrogen-Bond Rearrangements in the Halide–Dihydrate Complexes. *The journal of physical chemistry letters* **2019**, *10*, 2823-2828.

(5) Chaban, G. M.; Jung, J. O.; Gerber, R. B. Anharmonic Vibrational Spectroscopy of Hydrogen-Bonded Systems Directly Computed from ab Initio Potential Surfaces:(H2O) n, n= 2, 3; Cl-(H2O) n, n= 1, 2; H+ (H2O) n, n= 1, 2; H2O− CH3OH. *The Journal of Physical Chemistry A* **2000**, *104*, 2772-2779.

(6) Choi, J.-H.; Kuwata, K. T.; Cao, Y.-B.; Okumura, M. Vibrational Spectroscopy of the Cl-(H2O) n Anionic Clusters, n= 1− 5. *The Journal of Physical Chemistry A* **1998**, *102*, 503-507.

(7) Diken, E. G.; Headrick, J. M.; Roscioli, J. R.; Bopp, J. C.; Johnson, M. A.; McCoy, A. B.; Huang, X.; Carter, S.; Bowman, J. M. Argon Predissociation Spectroscopy of the OH-⊙ H2O and Cl-⊙ H2O Complexes in the 1000− 1900 cm-1 Region: Intramolecular

Bending Transitions and the Search for the Shared-Proton Fundamental in the Hydroxide Monohydrate. *The Journal of Physical Chemistry A* **2005**, *109*, 571-575.

(8) Dorsett, H. E.; Watts, R. O.; Xantheas, S. S. Probing Temperature Effects on the Hydrogen Bonding Network of the Cl-(H2O) 2 Cluster. *The Journal of Physical Chemistry A* **1999**, *103*, 3351-3355.

(9) Kim, J.; Lee, H. M.; Suh, S. B.; Majumdar, D.; Kim, K. S. Comparative ab initio study of the structures, energetics and spectra of X−·(H 2 O) n= 1–4 [X= F, Cl, Br, I] clusters. *The Journal of Chemical Physics* **2000**, *113*, 5259-5272.

(10) Punyain, W.; Takahashi, K. Theoretical calculation of the vibrational state dependent photodetachment spectra of X− H 2 O, X= F, Cl, Br. *Physical Chemistry Chemical Physics* **2016**, *18*, 26970-26979.

(11) Rheinecker, J.; Bowman, J. M. The calculated infrared spectrum of Cl− H 2 O using a new full dimensional ab initio potential surface and dipole moment surface. *The Journal of chemical physics* **2006**, *125*, 133206.

(12) Robertson, W. H.; Johnson, M. A. Molecular aspects of halide ion hydration: The cluster approach. *Annual review of physical chemistry* **2003**, *54*, 173-213.

(13) Roscioli, J. R.; Diken, E. G.; Johnson, M. A.; Horvath, S.; McCoy, A. B. Prying Apart a Water Molecule with Anionic H-Bonding: A Comparative Spectroscopic Study of the X-⊙ H2O (X= OH, O, F, Cl, and Br) Binary Complexes in the 600− 3800 cm-1 Region. *The Journal of Physical Chemistry A* **2006**, *110*, 4943-4952.

(14) Wang, X.-G.; Carrington Jr, T. Rovibrational levels and wavefunctions of Cl− H2O. *The Journal of Chemical Physics* **2014**, *140*, 204306.

(15) Xantheas, S. S. Quantitative Description of Hydrogen Bonding in Chloride− Water Clusters. *The Journal of Physical Chemistry* **1996**, *100*, 9703-9713.

(16) Zhao, H.; Xie, D.; Guo, H. Quantum dynamics of ClH2O− photodetachment: Isotope effect and impact of anion vibrational excitation. *The Journal of chemical physics* **2018**, *148*, 064305.

(17) Ayotte, P.; Nielsen, S. B.; Weddle, G. H.; Johnson, M. A.; Xantheas, S. S. Spectroscopic Observation of Ion-Induced Water Dimer Dissociation in the X-⊙(H2O) 2 (X= F, Cl, Br, I) Clusters. *The Journal of Physical Chemistry A* **1999**, *103*, 10665-10669.

(18) Horvath, S.; McCoy, A. B.; Elliott, B. M.; Weddle, G. H.; Roscioli, J. R.; Johnson, M. A. Anharmonicities and isotopic effects in the vibrational spectra of X−· H2O,· HDO, and· D2O [X= Cl, Br, and I] binary complexes. *The Journal of Physical Chemistry A* **2010**, *114*, 1556-1568.

(19) Galashev, A.; Sigon, F.; Servida, A. Molecular dynamic study of stabilization of the Cl−− Cl− anion pair in steam. *Journal of structural chemistry* **1996**, *37*, 252-259.

(20) Gao, J.; Boudon, S.; Wipff, G. Ab initio and crystal structure analysis of like-charged ion pairs. *Journal of the American Chemical Society* **1991**, *113*, 9610-9614.

(21) Choi, C. H.; Re, S.; Rashid, M. H.; Li, H.; Feig, M.; Sugita, Y. Solvent electronic polarization effects on Na+–Na+ and Cl—Cl–pair associations in aqueous solution. *The Journal of Physical Chemistry B* **2013**, *117*, 9273-9279.

(22) Ivanov, A. S.; Frenking, G.; Boldyrev, A. I. Stabilization of a Cl—Cl–Anion Pair in the Gas Phase: Ab Initio Microsolvation Study. *The Journal of Physical Chemistry A* **2014**, *118*, 7375-7384.

(23) Butchard, J. R.; Curnow, O. J.; Pipal, R. J.; Robinson, W. T.; Shang, R. Structural and spectroscopic investigations of non-planar tris (dialkylamino) cyclopropenium cations. *Journal of Physical Organic Chemistry* **2008**, *21*, 127-135.

(24) Butchard, J. R.; Curnow, O. J.; Garrett, D. J.; Maclagan, R. G.; Libowitzky, E.; Piccoli, P. M.; Schultz, A. J. Structural, theoretical and spectroscopic studies of the dichloride hexahydrate cube [Cl 2 (H 2 O) 6] 2−. *Dalton Transactions* **2012**, *41*, 11765-11775.

(25) Curnow, O. J.; Maclagan, R. G. Two Chloride Monohydrates Trapped in a Hydrophobic Pocket. *ChemPhysChem* **2012**, *13*, 3271-3274.

(26) Smith Jr, L. S.; Wertz, D. Solute structuring in aqueous lanthanum (III) chloride solutions. *Journal of the American Chemical Society* **1975**, *97*, 2365-2368.

(27) Copestake, A.; Neilson, G.; Enderby, J. The structure of a highly concentrated aqueous solution of lithium chloride. *Journal of Physics C: Solid State Physics* **1985**, *18*, 4211.

(28) Buckner, J. K.; Jorgensen, W. L. Energetics and hydration of the constituent ion pairs of tetramethylammonium chloride. *Journal of the American Chemical Society* **1989**, *111*, 2507-2516.

(29) Keasler, S. J.; Nellas, R. B.; Chen, B. Water mediated attraction between repulsive ions: A cluster-based simulation approach. *The Journal of chemical physics* **2006**, *125*, 144520.

(30) Dang, L. X.; Pettitt, B. M. Chloride ion pairs in water. *Journal of the American Chemical Society* **1987**, *109*, 5531-5532.

(31) Dang, L. X.; Pettitt, B. M. A theoretical study of like ion pairs in solution. *Journal of Physical Chemistry* **1990**, *94*, 4303-4308.

(32) Basu, A.; Das, G. Encapsulation of a discrete cyclic halide water tetramer [X 2 (H 2 O) 2] 2−, X= Cl−/Br− within a dimeric capsular assembly of a tripodal amide receptor. *Chemical Communications* **2013**, *49*, 3997-3999.

(33) Szumna, A.; Jurczak, J. Unusual Encapsulation of Two Anions in the Cavity of Neutral Macrocyclic Octalactam, Preliminary Communication. *Helvetica chimica acta* **2001**, *84*, 3760-3765.

(34) Wang, D.-X.; Fa, S.-X.; Liu, Y.; Hou, B.-Y.; Wang, M.-X. Anion-directed assembly of a rectangular supramolecular cage in the solid state with electron-deficient phenoxylated oxacalix [2] arene [2] triazine. *Chemical Communications* **2012**, *48*, 11458-11460.

(35) Safin, D. A.; Babashkina, M. G.; Robeyns, K.; Garcia, Y. Organic matrix-induced formation of a discrete cyclic [Cl 2 (H 2 O) 2] 2− cluster. *New Journal of Chemistry* **2017**, *41*, 8263-8269.

(36) Kopylovich, M. N.; Tronova, E. A.; Haukka, M.; Kirillov, A. M.; Kukushkin, V. Y.; Fraústo da Silva, J. J.; Pombeiro, A. J. Identification of hexameric water and hybrid water–chloride clusters intercalated in the crystal hosts of (imidoylamidine) nickel (II) complexes. *European Journal of Inorganic Chemistry* **2007**, *2007*, 4621-4627.

(37) Arunachalam, M.; Ghosh, P. A versatile tripodal amide receptor for the encapsulation of anions or hydrated anions via formation of dimeric capsules. *Inorganic chemistry* **2010**, *49*, 943-951.

(38) Chakraborty, S.; Dutta, R.; Arunachalam, M.; Ghosh, P. Encapsulation of [X 2 (H 2 O) 4] 2−(X= F/Cl) clusters by pyridyl terminated tripodal amide receptor in aqueous medium: single crystal X-ray structural evidence. *Dalton Transactions* **2014**, *43*, 2061-2068.

(39) Li, B.; Li, X.; Sun, X.; Wang, N. Halide-Anion Water Clusters in Cucurbit [6] uril Supramolecular Systems. *Chinese Journal of Chemistry* **2016**, *34*, 1114-1120.

(40) Ghosh, A. K.; Ghoshal, D.; Ribas, J.; Mostafa, G.; Chaudhuri, N. R. Hydrogen-bonded assembly of water and chloride in a 3D supramolecular host. *Crystal growth & design* **2006**, *6*, 36-39.

(41) Chen, W.-j.; Long, L.-S.; Huang, R.-B.; Zheng, L.-S. A Dihalide–Decahydrate Cluster of $[X_2(H_2O)_{10}]^{2-}$in a Supramolecular Architecture of $\{[Na_2(H2O)_6(H_2O@TMEQ[6])]\cdot 2\ (C_6H_5NO_3)\}X_2(H_2O)_{10}$(TMEQ[6]=α,α′,δ, δ′-Tetramethylcucurbit[6]uril; X= Cl, Br). *Crystal growth & design* **2013**, *13*, 2507-2513.

(42) Sibaev, M.; Crittenden, D. L. Quadratic corrections to harmonic vibrational frequencies outperform linear models. *The Journal of Physical Chemistry A* **2015**, *119*, 13107-13112.

(43) Wallace, A. J.; Jayasinghe, C. D.; Polson, M. I. J.; Curnow, O. J.; Crittenden, D. L. Cyclopropenium Cations Break the Rules of Attraction to Form Closely Bound Dimers. *Journal of the American Chemical Society* **2015**, *137*, 15528-15532.

(44) Cockroft, S. L.; Perkins, J.; Zonta, C.; Adams, H.; Spey, S. E.; Low, C. M. R.; Vinter, J. G.; Lawson, K. R.; Urch, C. J.; Hunter, C. A. Substituent effects on aromatic stacking interactions. *Organic & biomolecular chemistry* **2007**, *5*, 1062.

(45) Ringer, A. L.; Sherrill, C. D. Substituent Effects in Sandwich Configurations of Multiply Substituted Benzene Dimers Are Not Solely Governed By Electrostatic Control. *Journal of the American Chemical Society* **2009**, *131*, 4574-4575.

(46) Wheeler, S. E.; Houk, K. N. Substituent Effects in the Benzene Dimer are Due to Direct Interactions of the Substituents with the Unsubstituted Benzene. *Journal of the American Chemical Society* **2008**, *130*, 10854-10855.

(47) Peccati, F.; Laplaza, R. n.; Contreras-García, J. Overcoming Distrust in Solid State Simulations: Adding Error Bars to Computational Data. *Journal of physical chemistry. C* **2019**, *123*, 4767-4772.

Chapter 5

Characterisation of a Series of Discrete Dichloride Dihydrates and the Effect of Symmetry

---

"The ceaseless activity of their own inherent nature makes these stages moments of an organic unity, where they not merely do not contradict one another, but where one is as necessary as the other; and constitutes thereby the life of the whole."

George Wilhelm Friedrich Hegel

The material contained within this chapter will be published in a peer-reviewed journal article. The authors of this article and its title are stated below.

- Senthooran, R; Curnow, OJ; Brenner, T; Weiss, R; Ferreras, M; Crittenden, D; Characterisation of a Series of Discrete Dichloride Dihydrates and the Effect of Symmetry.

The contributions of each author are the following. Rathiga Senthoraan and Thomas Brenner synthesised and characterised the materials. Deborah Crittenden and Owen Curnow wrote the manuscript, and supervised the project alongside Robert Weiss. Manuel Ferreras Moreno was responsible for the computational design with Deborah Crittenden, and running the simulations.

Dr. Matthew I. J. Polson (University of Canterbury) is thanked for his assistance with the X-ray crystallography.

5.1. Introduction

Although there is good evidence for the formation of chloride-chloride ion pairs in aqueous solution, [1-10] the number of isolated discrete dichloride hydrate species $[Cl_2(H_2O)_n]^{2-}$ is very limited. We have recently reported examples of the monohydrate and tetrahydrate,[11,12] as well as having reported a hexahydrate several years ago.[13,14] There are a few examples of dihydrate clusters, and two of these might be described as discrete: Kleinman reported the first example of a discrete $[Cl_2(H_2O)_2]^{2-}$ cluster in 1992 with a fluorinated oxo-benzoquinolizine[15] and Wang *et al*. isolated one in 2012, however, the authors state that it exhibits strong Cl^-–π and H_2O–π interactions with electron-deficient regions of the triazine rings.[16] Basu et al. synthesized a dihydrate with an amide receptor in which the structure is stabilized by two NH---OH_2 and two NH---Cl^-hydrogen bonds.[17] Similarly, a dihydrate was reported by Szumna et al. as a co-crystal of a neutral octalactam receptor with tetrabutylammonium chloride in which the cluster lies in a monocyclic cavity held via four NH---OH_2 and four NH---Cl^- hydrogen bonds.[18] Safin et al. claimed a discrete dichloride dihydrate, however, the cluster is in fact supported by four amino-NH proton donor groups and in our opinion cannot be described as "discrete".[19]

Chloride-water clusters are fundamentally-important motifs for understanding a wide range of physicochemical processes in nature. For example, they play a vital role in biological and geological systems; particularly at water/membrane interfaces and in the mobility of ions through ion channels, as well as in electrical phenomena in the troposphere and ionosphere.[20-22] Indeed, there are extensive theoretical studies on monochloride hydrate $[Cl(H_2O)_n]^-$ clusters over the last several years.[23-35] The vibrational modes of a limited number of monochloride hydrates have been observed via Ar-tagging at low temperature in the gas phase.[36,37] On the other hand, there are only five reported ab initio studies on dichloride hydrates $[Cl_2(H_2O)_n]^{2-}$. [1-4,13,14,38]

In our work, we have been using triaminocyclopropenium (TAC) cations to isolate discrete anion species. TAC cations have a high-lying HOMO and their interactions with anions are consequently remarkably weak.[39-42] Herein, we present three examples of discrete dichloride dihydrates that each display a different symmetry in the crystalline environment and so allow a useful comparison between their infrared absorption spectra.

5.2. Methods

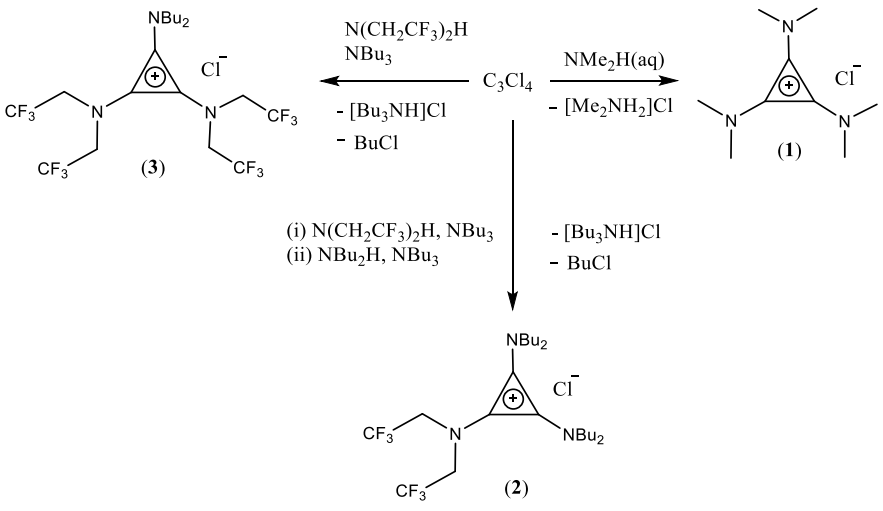
(Note that only computational methods are presented in this section. Experimental methods can be found in Appendix D)

An iterative optimization process was performed, keeping the carbon atom positions fixed throughout and alternately optimizing the positions of the attached hydrogen and fluorine atoms and then all atoms within the dichloride dihydrate. This produced an optimized geometry for the dichloride dihydrate within a fully relaxed environment subject only to the constraint that the carbon atoms remained in their crystallographically determined positions. A partial hessian calculation was carried out on the dichloride dihydrate[43,44] and all resultant harmonic frequencies were real, confirming that it represents a minimum on the potential energy surface subject to the applied constraints on surrounding carbon atom positions. An empirical quadratic correction model was applied to obtain predicted anharmonic vibrational frequencies.[45] All calculations were performed using QChem 4.2.[46]

5.3. Results and Discussion

5.3.1. Synthesis

The tris(dimethylamino)cyclopropenium chloride salt $[C_3(NMe_2)_3]Cl$ (**1**) has been known for some time.[39] It is readily prepared by reaction of dimethylamine with either C_3Cl_4 or C_3Cl_5H, however, we now prepare it using an aqueous solution of dimethylamine which is significantly safer than using the dry gas. The anhydrous salt is highly hygroscopic and even the monohydrate is very hygroscopic. Salt **1**.H_2O can be crystallized from CH_2Cl_2/diethylether at low temperature in the presence of traces of water. To date, the anhydrous salt has not been successfully crystallized. During our studies on fluorinated cyclopropenium cations, we synthesized the new TAC cation $[C_3(N(CH_2CF_3)_2)(NBu_2)_2]^+$ as the chloride salt **2** via the addition of a mixture of $HN(CH_2CF_3)_2$ and NBu_3 to tetrachlorocyclopropene followed by addition of NBu_2H and NBu_3 (Scheme 1). $[C_3(N(CH_2CF_3)_2)_2(NBu_2)]Cl$ (**3**), with two bis(trifluoroethyl)amino groups and only one dibutylamino group is prepared by the addition of a mixture of $HN(CH_2CF_3)_2$ and NBu_3 to tetrachlorocyclopropene, without additional NBu_2H. The salts were then crystallized as the monohydrates.



Scheme 1. Syntheses of $[C_3(NMe_2)_3]Cl$, $[C_3(N(CH_2CF_3)_2)_2(NBu_2)]Cl$ and $[C_3(N(CH_2CF_3)_2)(NBu_2)_2]Cl$.

5.3.2. Solid State Structures

$[C_3(NMe_2)_3]Cl.H_2O$ (**1**.H_2O) crystallizes in the monoclinic space group *I*2/*m* (Table 1) in which the asymmetric unit consists of a half each of the cation, chloride ion and water solvate. The cations form weak staggered dimers of parallel planes with a centroid-centroid distance of 3.1967(19) Å (Figure 1). Figure 2 shows the atomic labelling scheme. We have reported a number of TAC dimers previously with centroid-centroid distances varying from 3.20 to 3.35 Å.[47,48] The cation-cation distance is short, compared to the distance between neutral planes of graphite at 3.35 Å, for example, and is due to strong dispersion forces that outweigh the (highly-delocalized) charge repulsion. In each case, the cation adopts an umbrella conformation with the substituents bent away from the dimer. In the double salt of $[C_3(NMe_2)_3]Cl$ and $[NMe_2H_2]Cl$, half of the TAC cations were found to form dimers with bent TAC cations while the other half form monomers with flat TAC cations.[48] In **1**.H_2O, a dichloride dihydrate lies between each cationic dimer. The $[Cl_2(H_2O)_2]^{2-}$ cluster plane is at a slight angle to the TAC planes of 3.62(10)°, and the cluster centroid to TAC plane distances are 3.4222(15) Å. The centroid-centroid distances are 3.4876(9) Å, and they are offset by 0.672(6) Å. The cluster has crystallographic D_{2h} symmetry, so the Cl---O distances are all 3.2095(9) Å and the waters are equivalent. All of the H atom positions were refined: O–H = 0.854(18) Å and the Cl–H-O angle is effectively linear at 179.2(17)° with Cl---H = 2.356(18) Å. The overall cluster shape is a parallelogram with the angles at Cl and O significantly different from 90°: Cl–O–Cl = 108.72(4)° and O–Cl–O = 71.28(4)°.

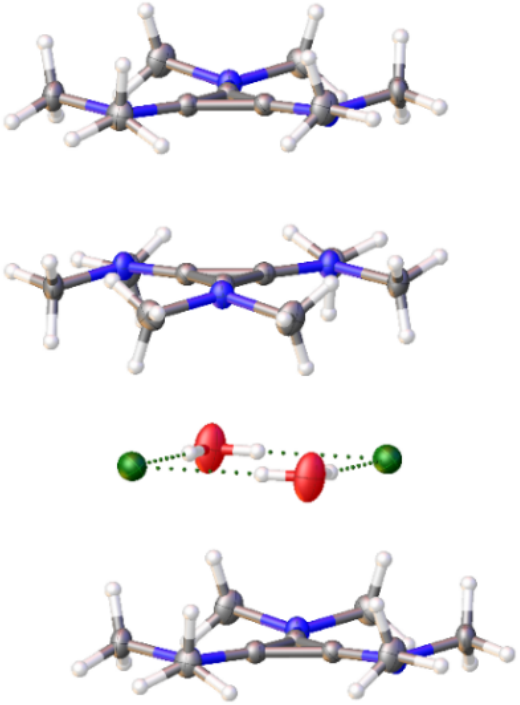


Figure 1. The geometry of $[C_3(NMe_2)_3]Cl.H_2O$ (**1**.H_2O) to illustrate the TAC dimer and its relationship with the dichloride dihydrate cluster.

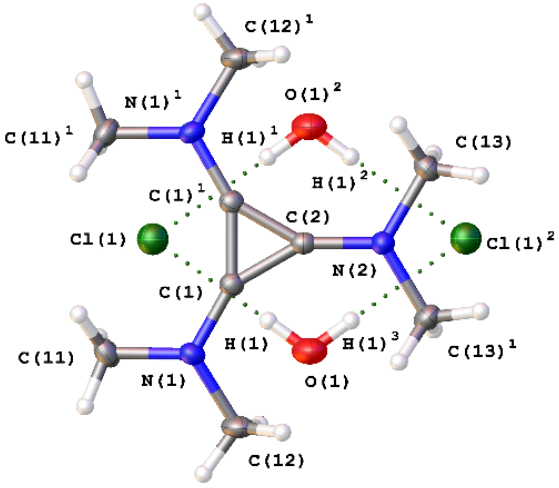


Figure 2. The atomic labelling scheme for **1**.H_2O.

$[C_3(N(CH_2CF_3)_2)(NBu_2)_2]Cl.H_2O$ (**2**.H_2O) forms in the triclinic space group *P*–1 with one each of the cation, chloride and water in the asymmetric unit. The chloride and water form a dichloride dihydrate cluster that is sandwiched by two cations via non-classical hydrogen bonds with the methylene groups of the trifluoroethyl substituents (Figure 3). Figure 4 shows the atomic labelling scheme. The cations have adopted a conformation in which the trifluoroethyl groups are on the same side of the TAC plane so as to maximize the hydrogen bonding with the dichloride hydrate cluster while the butyl groups alternate either side of the TAC plane. The dichloride hydrate cluster has crystallographic C_{2h} symmetry so, although the water molecules are equivalent, the Cl---O distances for each molecule are different (3.2182(15) Å for O1---Cl1 and 3.2568(18) Å for O1---Cl1^{1}). The H atoms were refined (O1–H1A = 0.85(4) and O1–H1B = 0.83(3) Å; H-O-H = 99(3)°) and the Cl–H-O angles found to be slightly bent

(O1-H1A-Cl1[1] = 176(3)° and O1-H1B-Cl1 = 170(3)°). Whereas the cluster in **1**.H_2O is a parallelogram, the cluster in **2**.H_2O is a slightly distorted square with very similar Cl---O distances and angles close to 90° (Cl–O–Cl = 90.26(4)°; O–Cl–O = 89.74(4)°).

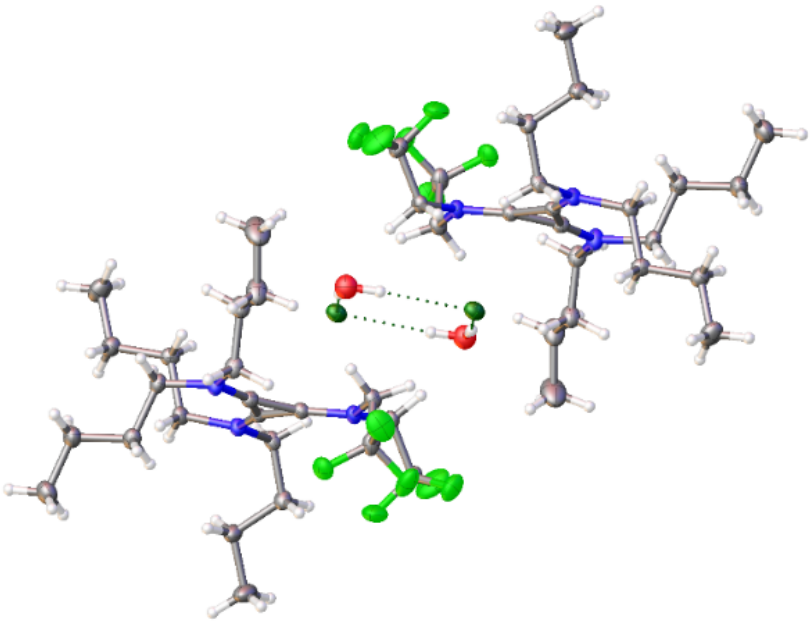


Figure 3. Two asymmetric units of $[C_3(N(CH_2CF_3)_2)(NBu_2)_2]Cl.H_2O$ (**2**.H_2O) to illustrate the relationship of the dichloride dihydrate cluster with the two closest cations.

The cluster has four CH–Cl hydrogen bonds (H---Cl = 2.66–3.03 Å) and one CH–O hydrogen bond (H---O = 2.5866(16) Å) to each cation above and below the plane and two CH–Cl hydrogen bonds (H---Cl = 2.83–3.04 Å) to each of two cations in the cluster plane. As expected, the hydrogen bonds to the trifluoroethyl methylene groups are noticeably shorter (stronger) than to the butyl groups.

Table 1. Structure refinement data for $[C_3(NMe_2)_3]Cl.H_2O$ (**1**.H_2O), $[C_3(N(CH_2CF_3)_2)(NBu_2)_2]Cl.H_2O$ (**2**.H_2O) and $[C_3(N(CH_2CF_3)_2)_2(NBu_2)]Cl.H_2O$ (**3**.H_2O).

	1.H_2O	**2**.H_2O	**3**.H_2O
formula	$C_9H_{20}ClN_3O$	$C_{23}H_{42}ClF_6N_3O$	$C_{19}H_{28}ClF_{12}N_3O$
a [Å]	9.3304(2)	10.0561(7)	23.8578(4)
b [Å]	13.6461(3)	11.0724(5)	10.9868(2)
c [Å]	10.1199(3)	13.5008(7)	19.6128(4)
α [°]	90	68.976(4)	90
β [°]	111.478(3)	83.661(5)	94.668(2)
γ [°]	90	84.001(5)	90
V [Å^3]	1199.03(6)	1391.18(14)	5123.87(16)
Z	4	2	8
ρ_{calc} [g cm^{-3}]	1.228	1.256	1.498
crystal system	monoclinic	triclinic	monoclinic
space group	$I2/m$	$P\bar{1}$	$I2/a$
shape/colour	needle/light brown	plate/colourless	plate/colourless
crystal size [mm]	0.201 × 0.143 × 0.126	0.214 × 0.063 × 0.034	0.155 × 0.129 × 0.059
μ [mm^{-1}]	2.631	1.746	2.305
T [K]	120.0(2)	120.01(10)	120.0(2)
$F(000)$	480	560	2368
2θ range [°]	11.046 – 154.144	7.036 – 154.102	7.436 – 154.418
index ranges	$-11 \le h \le 11$	$-12 \le h \le 10$	$-30 \le h \le 30$
	$-17 \le k \le 17$	$-13 \le k \le 11$	$-13 \le k \le 13$
	$-10 \le l \le 12$	$-16 \le l \le 17$	$-21 \le l \le 24$
reflections collected	12365	14525	24032
independent reflns	1322	5769	5373
R(int)	0.0266	0.0422	0.0284
data/restraints/parameters	1322/0/109	5769/0/319	5373/0/438
GoF on F^2	1.087	1.028	1.063
R_1/wR_2 [$I>2\sigma(I)$]	0.0262/0.0679	0.0397/0.0923	0.0377/0.0952
R_1/wR_2 (all data)	0.0266/0.0682	0.0586/0.0997	0.0416/0.0985
$\Delta\rho_{max/min}$/e [Å^{-3}]	0.218/–0.159	0.235/–0.378	0.439/–0.455
CCDC number	2003399	2003397	2003398

$[C_3(N(CH_2CF_3)_2)_2(NBu_2)]Cl.H_2O$ (**3**.H_2O) crystallizes in the monoclinic space group $I2/a$ (Table 1) in which the asymmetric unit consists of one cation, a chloride ion and two half-water solvates (Figure 5). The waters sit on a crystallographic C_2 axis so the dichloride dihydrate has crystallographic C_{2v} symmetry with equivalent chlorides but different waters (each with equivalent O–H bonds). Each cluster is sandwiched between two cations (Figure 5) with two additional cations that have very weak CH–O hydrogen bonds to one water. Furthermore, there are two cations (not shown in Figure 5) each weakly attached via two CH–Cl hydrogen bonds (Shown in Figure 4S). The atomic labelling scheme is shown in Figure 6.

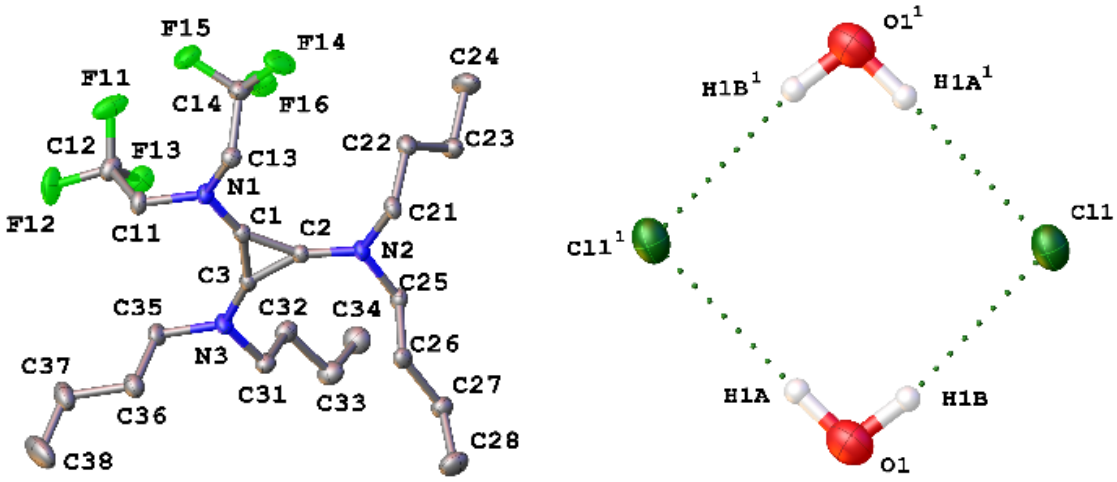


Figure 4. The atomic labelling scheme for **2**.H_2O

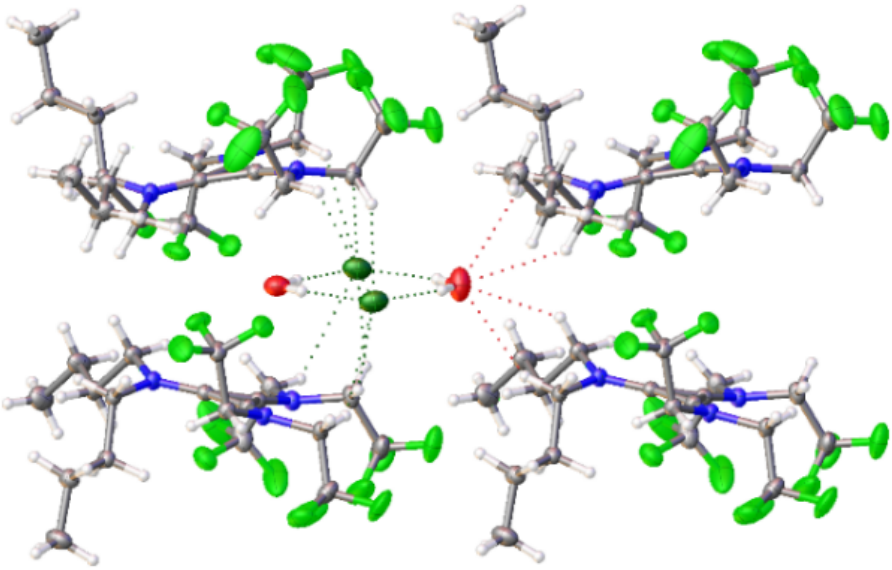


Figure 5. Side view of $[C_3(N(CH_2CF_3)_2)_2(NBu_2)]Cl.H_2O$ (**3**.H_2O) to show the main hydrogen-bonding interactions with the cluster.

The cations in **3**.H_2O each adopt a conformation in which all of the substituents are on the same side of the TAC cation plane except for one trifluoroethyl group. This arrangement appears to maximize the hydrogen bonding between the dichloride ion pair and the trifluoroethyl methylene groups which are hydrogen bonding exclusively to the chloride ions rather than the

waters. There are eight CH---Cl hydrogen bonds in the range 2.85–2.90 Å for the H---Cl distance and two shorter hydrogen bonds at 2.72(2) Å.

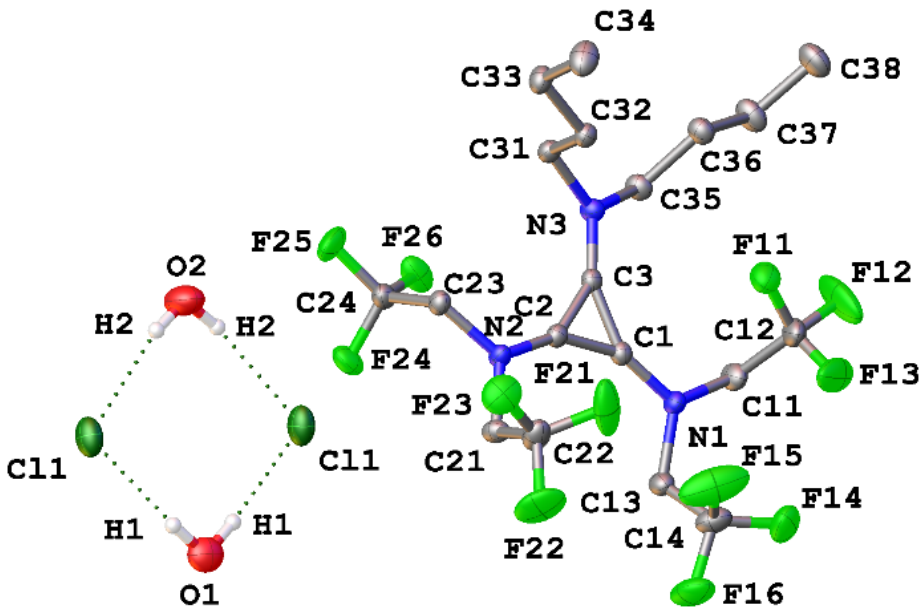


Figure 6. The atomic labelling scheme for **3**.H_2O.

The dichloride hydrate cluster in **3**.H_2O has one water at a similar distance to the chlorides as in **1**.H_2O Cl---O = (3.2051(15) Å versus 3.2095(9) Å, respectively) whereas the other water has a much greater Cl---O distance (3.2975(15) Å) than in any of the other clusters. It is this second water (O2) that has very weak CH–O hydrogen bonds to the alpha-methylene groups of the butyl substituents (H---O = 2.82(2) Å and 2.84(2) Å in **3**.H_2O compared to 2.5866(16) Å in **2**.H_2O and 2.603(19) Å and 2.613(17) Å in **1**.H_2O). Again, the O-H–Cl angles are bent (175(2)° and 168(2)° for O1 and O2, respectively).

Table 2 compares the $[Cl_2(H_2O)_2]^{2-}$ clusters described here with those reported earlier. Coincidentally, the other clusters all have crystallographic C_{2h} symmetry. Nonetheless, the structure reported by Safin et al. is remarkably similar to that in **1** with Cl–O distances within 0.03 Å and the angles at Cl and O within just 1°. The structure reported by Basu has the shortest

average Cl–O distances. Choi and co-workers recently reported a calculated structure for the dichloride dihydrate at the CCSD(T)/aug-cc-pVTZ//MP2/aug-cc-pVTZ level: The calculated Cl---O and Cl---Cl distances (3.38 Å and 5.53 Å, respectively) are greater than is observed in any of the solid state structures.[3] The O---O distance is in the middle of the experimental range at 3.89 Å; Interestingly, the O-Cl-O and Cl-O-Cl angles (70.2° and 109.8°, respectively) are very similar to that of **1**.H_2O and Safin's structure. Our expectation would be that, of all the structures, **1**.H_2O would be the most discrete example due to the expected weaker hydrogen bonding to the surrounding cations. The calculated structure tends to support that: although the calculated Cl---O distances (attractive hydrogen bonding) and the Cl---Cl distance (electrostatic repulsion) are longer, it would be expected that the calculation "error" would be consistently short or long and so the resultant angles might be correct.

Table 2. Structural parameters for $[Cl_2(H_2O)_2]^{2-}$ clusters.

Parameter	**Calculated**[a]	**1**.H_2O	**2**.H_2O	**3**.H_2O	Basu[17]	Wang[16]	Szumna[18]	Safin[19]	Kleinman[15]
O1---O2 [Å]	3.89	3.740(3)	4.568(3)	4.510(3)	4.286(5)	3.492(3)	4.410(15)	3.744(4)	4.191(7)
Cl1---Cl2 [Å]	5.53	5.2166(8)	4.5888(9)	4.6830(8)	4.623(2)	5.326(2)	4.875(4)	5.1688(16)	4.790(3)
Cl1---O1 [Å]	3.38	3.2095(9)	3.2568(18)	3.2051(15)	3.127(2)	3.123(2)	3.268(7)	3.180(2)	3.173(4)
Cl2---O1 [Å]	3.38	3.2095(9)	3.2182(15)	3.2051(15)	3.177(2)	3.245(2)	3.306(9)	3.203(2)	3.192(4)
Cl1---O2 [Å]	3.38	3.2095(9)	3.2182(15)	3.2975(15)	3.177(2)	3.245(2)	3.306(9)	3.203(2)	3.192(4)
Cl2---O2 [Å]	3.38	3.2095(9)	3.2568(18)	3.2975(15)	3.127(2)	3.123(2)	3.268(7)	3.180(2)	3.173(4)
Cl1–O1–Cl2 [°]	109.8	108.72(4)	90.26(4)	93.87(6)	94.33(7)	113.51(5)	95.7(2)	108.16(6)	97.64(10)
Cl1–O2–Cl2 [°]	109.8	108.72(4)	90.26(4)	90.48(5)	94.33(7)	113.51(5)	95.7(2)	108.16(6)	97.64(10)
O1–Cl1–O2 [°]	70.2	71.28(4)	89.74(4)	87.83(4)	85.67(7)	66.49(5)	84.3(2)	71.84(6)	82.36(10)
O1–Cl2–O2 [°]	70.2	71.28(4)	89.74(4)	87.83(4)	85.67(7)	66.49(5)	84.3(2)	71.84(6)	82.36(10)

[a] CCSD(T)/aug-cc-pVTZ//MP2/aug-cc-pVTZ level.[3]

5.3.3. Infrared spectroscopy

Vibrational spectroscopy provides detailed information about the bonding within and around solvated ion complexes.[49] We have collected infrared spectra on the three salts

described here as well as deuterated and partially-deuterated samples. The results are summarised in Tables 3 and 4.

Table 3. Experimental and calculated vibrational frequencies (in cm^{-1}) for $[Cl_2(H_2O)_2]^{2-}$ clusters.

1.H_2O	**2**.H_2O	**3**.H_2O	Calc[a]	Assignment
3520	3560	3560		Combination band
3427	3444	3475	3515	ν(asym,syn)
		3426	3499	ν(asym,anti)
		3373	3426	ν(sym,syn)
3368	3392	3373	3422	ν(sym,anti)
		3281	3362[b]	2ν(bend,syn)
3234	3270	3267	3326[b]	2ν(bend,anti)
		–	1681	ν(bend,syn)
1617	1645	1645	1663	ν(bend,anti)

[a] B3LYP/6-31+G* for C_{2v} symmetric complex, with empirical anharmonic scaling correction applied[45] ($\Delta_{anh} = 0.011214\nu_h + 0.000010982\nu_h^2$)
[b] Bending overtones estimated as double the bending fundamentals

Firstly, it should be noted that the average Cl---O distances in **1**.H_2O, **2**.H_2O and **3**.H_2O increases from 3.210 to 3.238 to 3.251 Å, respectively, and therefore the O–H bond strengths and stretching frequencies should increase in that order. However, not all bands are IR active, depending on the symmetry.

The dichloride dihydrate of **1**.H_2O has the highest symmetry, D_{2h}, and would be expected to have three fundamental IR-active bands in the water bending and stretching region (Figure 7). The stretching bands are labelled sym or asym for the symmetric and asymmetric water stretches, respectively, syn if there is a mirror plane through the Cl–Cl axis, and anti if there is no mirror plane. This is also the case for C_{2h}-symmetric cluster in **2**.H_2O. These IR-active bands appear at 1617 cm^{-1} (bend,anti), 3368 cm^{-1} (sym,anti) and 3427 cm^{-1} (asym,syn) for **1**.H_2O (Figure 8) and at slightly higher energy for **2**.H_2O (1645 cm^{-1}, 3392 cm^{-1} and 3444 cm^{-1}, respectively), which is consistent with the longer Cl---O distances in **2**.H_2O. Bending overtones

appear in the expected positions (3234 and 3270 cm^{-1} for **1**.H_2O and **2**.H_2O, respectively) and there is also a higher energy shoulder at 3520 cm^{-1} for **1**.H_2O which is likely due to a combination band of a stretching mode and a chloride-water librational mode.

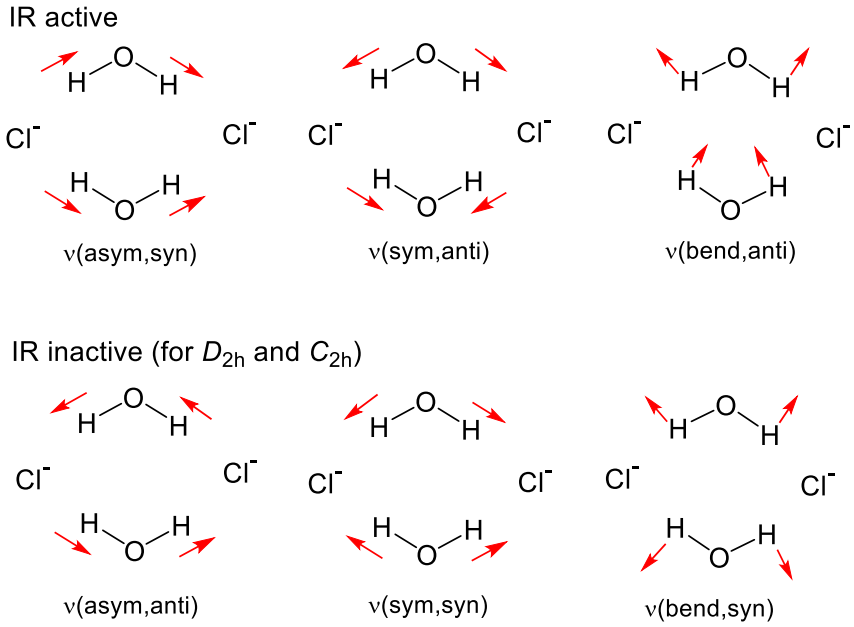


Figure 7. Fundamental water stretching and bending modes for $[Cl_2(H_2O_2]^{2-}$ clusters.

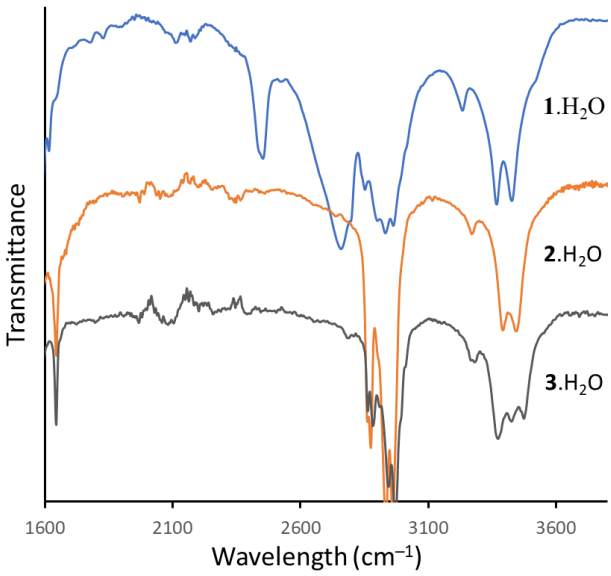


Figure 8. FT-IR spectra of **1**.H_2O, **2**.H_2O and **3**.H_2O.

The IR spectrum of **3**.H_2O, which has a C_{2v}-symmetric cluster, is more complex since all six of the vibrational modes shown in Figure 7 are now IR active. Although there is only one bending mode observed, bending modes typically do not vary greatly in energy compared to stretching modes. Nonetheless, the bending overtone is split into two bands of approximately equal intensity at 3281 and 3267 cm^{-1}. In the stretching region, we now see three strong bands, at 3475, 3426, and 3373 cm^{-1}. There is also a weak high-energy combination band at 3560 cm^{-1}.

In order to assist our assignment of the IR bands, we calculated the structure and vibrational frequencies for a dichloride dihydrate with di- and tri-fluoromethane molecules in fixed positions to mimic the environment of the cluster in **3**.H_2O. This produced a C_{2v}-symmetric cluster with a small difference between the Cl–water distances (see Appendix D). There are four distinct stretching vibrational modes and two bending modes (Table 3 and Figure 7). Although agreement with experiment is only qualitative due to both the level of *ab initio* theory employed, and the truncated nature of the model system, these results nonetheless allow the nature of the vibrations to be unambiguously assigned.

Based upon the fact that the empirical anharmonicity correction scheme used is designed to systematically overestimate predicted fundamentals (i.e. provide an upper bound to the experimentally-observed values), and taking symmetry splitting patterns into account, the two highest energy stretching bands are assigned to the asymmetric modes ν(asym,anti) < ν(asym,syn) while the lowest energy band at 3373 cm^{-1} is assigned to coincident symmetric modes ν(sym,anti) and ν(sym,syn) based on the smaller calculated difference and the greater intensity of the band. Geometrically, the change in dipole moment along the ν(sym,syn) mode will be small, and therefore this mode is expected to make only a small contribution to the intensity of this band.

For **2**.H_2O and **3**.H_2O, two sets of deuterated samples were prepared and their infrared spectra collected (Figures 9 and 10, respectively). In one set, denoted **2**.D_2O and **3**.D_2O, the samples were almost completely deuterated with only a small quantity of residual H present. In the other set (**2**.HDO and **3**.HDO), the samples were formulated with equal quantities of D_2O and H_2O, which randomize to form a statistical mixture of H_2O, D_2O and HDO molecules.

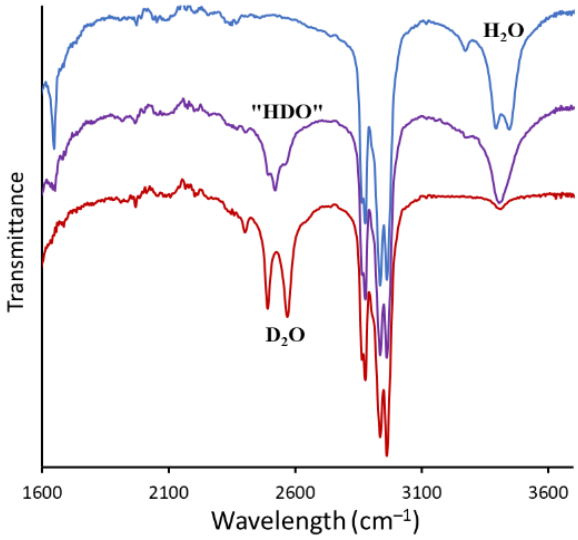


Figure 9. FT-IR spectra of **2**.H_2O, **2**.HDO and **2**.D_2O.

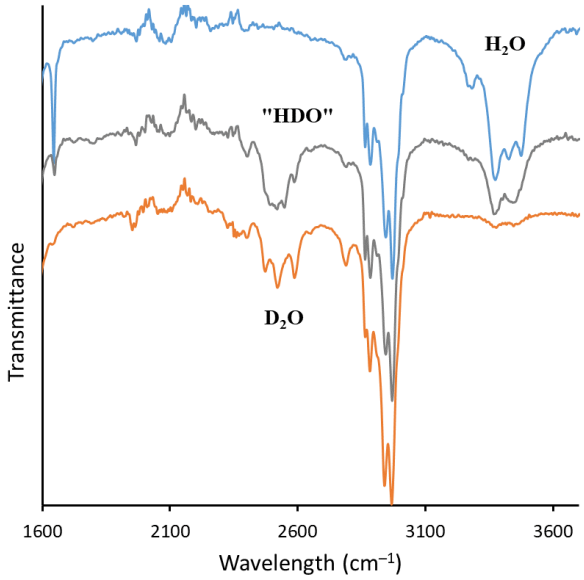


Figure 10. FT-IR spectra of **3**.H_2O, **3**.HDO and **3**.D_2O

The dominant stretching bands in the **2**.D_2O and **3**.D_2O spectra can be straightforwardly assigned by analogy with the fully hydrogenated species, as listed in Table 4. However, for completeness, we also report computational predictions for these species. The weak bending overtone and combination bands are also present.

Table 4. Experimental vibrational frequencies (in cm^{-1}) for fully deuterated samples of **2** and **3** and calculated vibrational frequencies for a truncated model of **3**.

2.D_2O	**3**.D_2O	Calc[a]	Assignment
2640	2650		Combination band
2568	2588	2611	ν(asym,syn)
	2520	2598	ν(asym,anti)
	2474	2502	ν(sym,syn)
2490	2474	2499	ν(sym,anti)
2400	2400	2443[b]	2ν(DOD)

[a] B3LYP/6-31+G* for C_{2v} symmetric complex, with empirical anharmonic scaling correction applied[21] ($\Delta_{anh} = 0.011214\nu_h + 0.000010982\nu_h^2$)
[b] Bending overtones estimated as double the bending fundamentals

The trace amount of H in the "fully deuterated" samples will be largely present as $[Cl_2(D_2O)(HDO)]^{2-}$ with only very small amounts of the dideuterated isomers $[Cl_2(D_2O)(H_2O)]^{2-}$ and $[Cl_2(HDO)_2]^{2-}$. Thus, the weak band at 3410 cm^{-1} in the spectrum of **2**.D_2O can be assigned to the O–H stretch of $[Cl_2(D_2O)(HDO)]^{2-}$ which, as expected, lies in energy between the symmetric and asymmetric bands for **2**.H_2O.

The samples of **2**.HDO and **3**.HDO contain a statistical mixture of five isotopomers in a 1:4:6:4:1 ratio, and some of these also have isomers, which complicates the interpretation of the corresponding spectra. To assist in assigning these spectra, we have computed IR band centres for the most abundant isotopomers of **3**.HDO; $[Cl_2(D_2O)(HDO)]^{2-}$, $[Cl_2(H_2O)(HDO)]^{2-}$, $[Cl_2(HDO)_2]^{2-}$ and $[Cl_2(H_2O)(D_2O)]^{2-}$. Predicted and observed band centres are listed in Table 5.

Table 5. Experimental vibrational frequencies (in cm^{-1}) for partially-deuterated samples of **2** and **3** and calculated vibrational frequencies for isotopomers of **3**, each with two possible isomers that give rise to very similar IR stretching fundamentals (see supplementary information for more details).

Experimental		Calculated[a]				Assignment
2."HDO"	**3**."HDO"	(D_2O) (HDO)[b]	(H_2O) (HDO)[b]	$(HDO)_2$[c]	(H_2O) (D_2O)[d]	
	3460		3507		3507	H_2O asym str
3410		3467	3466	3462,3472		O-H stretch
	3377		3424		3424	H_2O sym str
2557		2605			2605	D_2O asym str
2519	n.r.[e]	2551	2551	2548,2555		O-D stretch
2492		2501			2501	D_2O sym str

[a] B3LYP/6-31+G* for C_{2v} symmetric complex, with empirical anharmonic scaling correction applied[45]($\Delta_{anh} = 0.011214\nu_h + 0.000010982\nu_h^2$)

[b] Each complex forms 4 symmetry-distinct isotopomers, but symmetry splittings are very small. Therefore, averaged predicted band centres are reported.

[c] Complex forms 2 symmetry-distinct isotopomers, each with two coupled modes, as illustrated in Figure 4. All 4 predicted band centres are reported.

[d] Complex forms 2 symmetry-distinct isotopomers, but symmetry splittings are very small. Therefore, averaged predicted band centres are reported.

[e] Band centre is not resolved (n.r.)

For **2**.HDO, two sharp, high intensity bands dominate the O-H and O-D stretching regions of the spectrum. They lie intermediate in energy between the symmetric and asymmetric stretching bands of H_2O and D_2O, and so can be assigned to the O–H and O–D stretching modes that are statistically expected to be highest in intensity. Band centres are predicted to occur within narrow ranges, even for the different isotopomers of $[Cl_2(HDO)_2]^{2-}$ illustrated in Figure 11. Three of the four modes are IR active for each of O–H and O–D. The mono- and tri-deuterated isotopomers $[Cl_2(H_2O)(HDO)]^{2-}$ and $[Cl_2(D_2O)(HDO)]^{2-}$ also contribute to the observed O-H and O-D stretching intensities.

In the O–D region, shoulders on the high intensity O–D stretching band are observed. They are assigned to D_2O symmetric and asymmetric stretching modes contributed by the dideuterated isotopomer $[Cl_2(D_2O)(H_2O)]^{2-}$ and the tri-deuterated isotopomer $[Cl_2(D_2O)(HDO)]^{2-}$ in a 2:1 ratio.

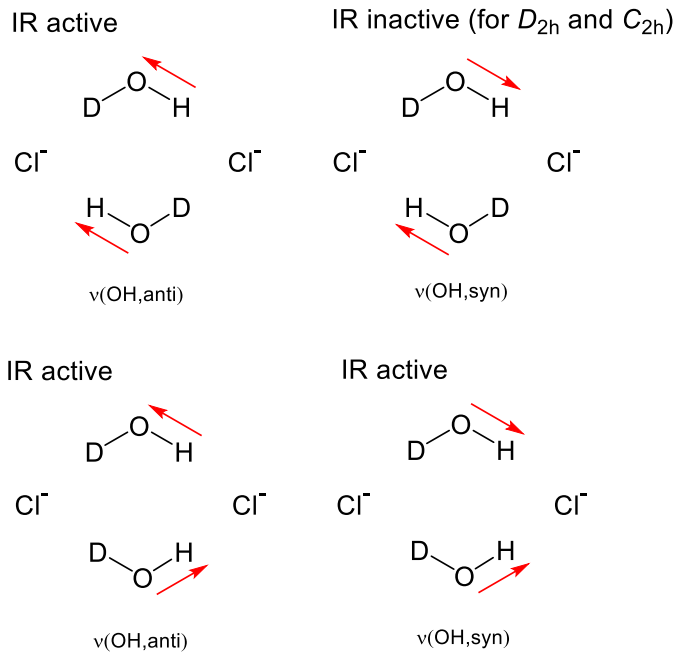


Figure 11. Coupled O-H stretching modes of $[Cl_2(HDO)_2]^{2-}$. O-D modes are defined analogously.

In the O–H region, the mono-deuterated isotopomer $[Cl_2(H_2O)(HDO)]^{2-}$ would contribute instead, but these shoulders cannot be clearly resolved because the sym/asym splitting is smaller in H_2O than in D_2O and so there is a greater degree of overlap between the H_2O and O–H stretching bands.

The IR spectrum of **3**.HDO is more complex. Its interpretation further complicated by the fact that all bands are now IR active due to the lower symmetry of the complex, in addition to contributions from multiple isotopomers and multiple isomers. The O–D region is complex and there no obvious bands that can be resolved and assigned. In the O–H region, on the other hand, the clearest change lies in the appearance of a strong new band at 3460 cm^{-1}, close to the centre of the ν(asym,syn) and ν(asym,anti) bands from the **3**.H_2O spectrum, while the symmetric H_2O stretching band at 3377 cm^{-1} remains unchanged. We note that this band cannot be resolved into symmetry-split components, even in the undeuterated system, **3**.H_2O. We therefore assign the OH bands to the asymmetric and symmetric stretches of H_2O while

multiple separate O-H stretching bands originating from the HDO fragment cannot be clearly resolved.

Finally, it is worth noting that symmetry-splitting effects arising from the lower symmetry of complex **3** are much smaller than the effects of isotopic substitution. Therefore, the differences between the **2**.HDO and **3**.H_2O spectra are not due to the difference in symmetry between the two species, but rather the difference in isotopomer contributions.

5.4. Conclusions

We have isolated and characterized three new "discrete" dichloride dihydrate clusters, each with different crystallographically-imposed symmetry and different weak non-classical hydrogen bonding interactions with the cations that lead to subtle structural differences and dramatic spectroscopic differences. Most notably, the C_{2v}-symmetric cluster in **3**.H_2O exhibits three stretching bands rather than two, and the bending overtone is split. Separate syn and anti bands (and overtone splitting) are observed due to coupling between the vibrational modes of the symmetry inequivalent water molecules. With symmetry-equivalent waters one of these bands is IR inactive. Computational results were used to guide band assignments and differentiate between symmetry-splitting and isotope substitution effects.

References

(1) Galashev, A.; Sigon, F.; Servida, A. Molecular dynamic study of stabilization of the Cl−− Cl− anion pair in steam. *Journal of structural chemistry* **1996**, *37*, 252-259.

(2) Gao, J.; Boudon, S.; Wipff, G. Ab initio and crystal structure analysis of like-charged ion pairs. *Journal of the American Chemical Society* **1991**, *113*, 9610-9614.

(3) Choi, C. H.; Re, S.; Rashid, M. H.; Li, H.; Feig, M.; Sugita, Y. Solvent electronic polarization effects on Na+–Na+ and Cl—Cl–pair associations in aqueous solution. *The Journal of Physical Chemistry B* **2013**, *117*, 9273-9279.

(4) Ivanov, A. S.; Frenking, G.; Boldyrev, A. I. Stabilization of a Cl—Cl–Anion Pair in the Gas Phase: Ab Initio Microsolvation Study. *The Journal of Physical Chemistry A* **2014**, *118*, 7375-7384.

(5) Smith Jr, L. S.; Wertz, D. Solute structuring in aqueous lanthanum (III) chloride solutions. *Journal of the American Chemical Society* **1975**, *97*, 2365-2368.

(6) Copestake, A.; Neilson, G.; Enderby, J. The structure of a highly concentrated aqueous solution of lithium chloride. *Journal of Physics C: Solid State Physics* **1985**, *18*, 4211.

(7) Buckner, J. K.; Jorgensen, W. L. Energetics and hydration of the constituent ion pairs of tetramethylammonium chloride. *Journal of the American Chemical Society* **1989**, *111*, 2507-2516.

(8) Keasler, S. J.; Nellas, R. B.; Chen, B. Water mediated attraction between repulsive ions: A cluster-based simulation approach. *The Journal of chemical physics* **2006**, *125*, 144520.

(9) Dang, L. X.; Pettitt, B. M. Chloride ion pairs in water. *Journal of the American Chemical Society* **1987**, *109*, 5531-5532.

(10) Dang, L. X.; Pettitt, B. M. A theoretical study of like ion pairs in solution. *Journal of Physical Chemistry* **1990**, *94*, 4303-4308.

(11) Curnow, O. J.; Senthooran, R. One water to bind a chloride-chloride ion pair: isolation of discrete [Cl 2 (H2O)]2^{-} in the solid state. *Dalton Transactions,* **2020**, 49(28), 9579-9582.

(12) Abdelbassit, M. S.; Curnow, O. J.; Ferreras, M.; Crittenden, D. L. A Discrete Dichloride Tetrahydrate Trapped by a Cyclopropenium Cation: Structure and Spectroscopic Properties. *ChemPlusChem* **2020**, *85*, 927-932.

(13) Butchard, J. R.; Curnow, O. J.; Garrett, D. J.; Maclagan, R. G.; Libowitzky, E.; Piccoli, P. M.; Schultz, A. J. Structural, theoretical and spectroscopic studies of the dichloride hexahydrate cube [Cl 2 (H 2 O) 6] 2−. *Dalton Transactions* **2012**, *41*, 11765-11775.

(14) Butchard, J. R.; Curnow, O. J.; Pipal, R. J.; Robinson, W. T.; Shang, R. Structural and spectroscopic investigations of non-planar tris (dialkylamino) cyclopropenium cations. *Journal of Physical Organic Chemistry* **2008**, *21*, 127-135.

(15) Kleinman, E. F.; Bordner, J.; Newhouse, B. J.; MacFerrin, K. Striking example of a contact chloride ion pair with bridging water molecules in the solid state. *Journal of the American Chemical Society* **1992**, *114*, 4945-4946.

(16) Wang, D.-X.; Fa, S.-X.; Liu, Y.; Hou, B.-Y.; Wang, M.-X. Anion-directed assembly of a rectangular supramolecular cage in the solid state with electron-deficient phenoxylated oxacalix [2] arene [2] triazine. *Chemical Communications* **2012**, *48*, 11458-11460.

(17) Basu, A.; Das, G. Encapsulation of a discrete cyclic halide water tetramer [X 2 (H 2 O) 2] 2−, X= Cl−/Br− within a dimeric capsular assembly of a tripodal amide receptor. *Chemical Communications* **2013**, *49*, 3997-3999.

(18) Szumna, A.; Jurczak, J. Unusual Encapsulation of Two Anions in the Cavity of Neutral Macrocyclic Octalactam, Preliminary Communication. *Helvetica chimica acta* **2001**, *84*, 3760-3765.

(19) Safin, D. A.; Babashkina, M. G.; Robeyns, K.; Garcia, Y. Organic matrix-induced formation of a discrete cyclic [Cl 2 (H 2 O) 2] 2− cluster. *New Journal of Chemistry* **2017**, *41*, 8263-8269.

(20) Castleman, A.; Bowen, K. Clusters: Structure, energetics, and dynamics of intermediate states of matter. *The Journal of Physical Chemistry* **1996**, *100*, 12911-12944.

(21) Ohtaki, H.; Radnai, T. Structure and dynamics of hydrated ions. *Chemical Reviews* **1993**, *93*, 1157-1204.

(22) Roux, B.; Bernèche, S.; Egwolf, B.; Lev, B.; Noskov, S. Y.; Rowley, C. N.; Yu, H. Ion selectivity in channels and transporters. *Journal of general physiology* **2011**, *137*, 415-426.

(23) Bajaj, P.; Zhuang, D.; Paesani, F. Specific Ion Effects on Hydrogen-Bond Rearrangements in the Halide–Dihydrate Complexes. *The journal of physical chemistry letters* **2019**, *10*, 2823-2828.

(24) Chaban, G. M.; Jung, J. O.; Gerber, R. B. Anharmonic Vibrational Spectroscopy of Hydrogen-Bonded Systems Directly Computed from ab Initio Potential Surfaces:(H2O) n, n= 2, 3; Cl-(H2O) n, n= 1, 2; H+ (H2O) n, n= 1, 2; H2O− CH3OH. *The Journal of Physical Chemistry A* **2000**, *104*, 2772-2779.

(25) Choi, J.-H.; Kuwata, K. T.; Cao, Y.-B.; Okumura, M. Vibrational Spectroscopy of the Cl-(H2O) n Anionic Clusters, n= 1− 5. *The Journal of Physical Chemistry A* **1998**, *102*, 503-507.

(26) Diken, E. G.; Headrick, J. M.; Roscioli, J. R.; Bopp, J. C.; Johnson, M. A.; McCoy, A. B.; Huang, X.; Carter, S.; Bowman, J. M. Argon Predissociation Spectroscopy of the OH-⊙ H2O and Cl-⊙ H2O Complexes in the 1000− 1900 cm-1 Region: Intramolecular

Bending Transitions and the Search for the Shared-Proton Fundamental in the Hydroxide Monohydrate. *The Journal of Physical Chemistry A* **2005**, *109*, 571-575.

(27) Dorsett, H. E.; Watts, R. O.; Xantheas, S. S. Probing Temperature Effects on the Hydrogen Bonding Network of the Cl-(H2O) 2 Cluster. *The Journal of Physical Chemistry A* **1999**, *103*, 3351-3355.

(28) Kim, J.; Lee, H. M.; Suh, S. B.; Majumdar, D.; Kim, K. S. Comparative ab initio study of the structures, energetics and spectra of X−·(H 2 O) n= 1–4 [X= F, Cl, Br, I] clusters. *The Journal of Chemical Physics* **2000**, *113*, 5259-5272.

(29) Punyain, W.; Takahashi, K. Theoretical calculation of the vibrational state dependent photodetachment spectra of X− H 2 O, X= F, Cl, Br. *Physical Chemistry Chemical Physics* **2016**, *18*, 26970-26979.

(30) Rheinecker, J.; Bowman, J. M. The calculated infrared spectrum of Cl− H 2 O using a new full dimensional ab initio potential surface and dipole moment surface. *The Journal of chemical physics* **2006**, *125*, 133206.

(31) Robertson, W. H.; Johnson, M. A. Molecular aspects of halide ion hydration: The cluster approach. *Annual review of physical chemistry* **2003**, *54*, 173-213.

(32) Roscioli, J. R.; Diken, E. G.; Johnson, M. A.; Horvath, S.; McCoy, A. B. Prying Apart a Water Molecule with Anionic H-Bonding: A Comparative Spectroscopic Study of the X-⊙ H2O (X= OH, O, F, Cl, and Br) Binary Complexes in the 600− 3800 cm-1 Region. *The Journal of Physical Chemistry A* **2006**, *110*, 4943-4952.

(33) Wang, X.-G.; Carrington Jr, T. Rovibrational levels and wavefunctions of Cl− H2O. *The Journal of Chemical Physics* **2014**, *140*, 204306.

(34) Xantheas, S. S. Quantitative Description of Hydrogen Bonding in Chloride− Water Clusters. *The Journal of Physical Chemistry* **1996**, *100*, 9703-9713.

(35) Zhao, H.; Xie, D.; Guo, H. Quantum dynamics of ClH2O− photodetachment: Isotope effect and impact of anion vibrational excitation. *The Journal of chemical physics* **2018**, *148*, 064305.

(36) Ayotte, P.; Nielsen, S. B.; Weddle, G. H.; Johnson, M. A.; Xantheas, S. S. Spectroscopic Observation of Ion-Induced Water Dimer Dissociation in the X-⊙(H2O) 2 (X= F, Cl, Br, I) Clusters. *The Journal of Physical Chemistry A* **1999**, *103*, 10665-10669.

(37) Horvath, S.; McCoy, A. B.; Elliott, B. M.; Weddle, G. H.; Roscioli, J. R.; Johnson, M. A. Anharmonicities and isotopic effects in the vibrational spectra of X−· H2O,· HDO, and· D2O [X= Cl, Br, and I] binary complexes. *The Journal of Physical Chemistry A* **2010**, *114*, 1556-1568.

(38) Curnow, O. J.; Maclagan, R. G. Two Chloride Monohydrates Trapped in a Hydrophobic Pocket. *ChemPhysChem* **2012**, *13*, 3271-3274.

(39) Weiss, R.; Schloter, K. Stable radical dications. *Tetrahedron Letters* **1975**, *16*, 3491-3494.

(40) Weiss, R.; Brenner, T.; Hampel, F.; Wolski, A. The Consequences of an Electrostatic "Forced Marriage" between Two Electron-Rich Particles: Strained Ion Pairs. *Angewandte Chemie International Edition in English* **1995**, *34*, 439-441.

(41) Weiss, R.; Rechinger, M.; Hampel, F.; Wolski, A. Stable 1: 1 adducts from iodoacetylenes and iodide ions: ion pair strain as an additional driving force? *Angewandte Chemie International Edition in English* **1995**, *34*, 441-443.

(42) Weiss, R.; Schwab, O.; Hampel, F. Ion-Pair Strain as the Driving Force for Hypervalent Adduct Formation between Iodide Ions and Substituted Iodobenzenes: Structural Alternatives to Meisenheimer Complexes. *Chemistry–A European Journal* **1999**, *5*, 968-974.

(43) Head, J. D. Computation of vibrational frequencies for adsorbates on surfaces. *International journal of quantum chemistry* **1997**, *65*, 827-838.

(44) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. Vibrational modes in partially optimized molecular systems. *The Journal of chemical physics* **2007**, *126*, 224102.

(45) Sibaev, M.; Crittenden, D. L. Quadratic corrections to harmonic vibrational frequencies outperform linear models. *The Journal of Physical Chemistry A* **2015**, *119*, 13107-13112.

(46) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics* **2015**, *113*, 184-215.

(47) Wallace, A. J.; Jayasinghe, C. D.; Polson, M. I. J.; Curnow, O. J.; Crittenden, D. L. Cyclopropenium Cations Break the Rules of Attraction to Form Closely Bound Dimers. *Journal of the American Chemical Society* **2015**, *137*, 15528-15532.

(48) Curnow, O. J.; Jayasinghe, C. D.; Polson, M. I.; Yunis, R. Triaminocyclopropenium Halide and Triiodide Salts: The Formation of Cyclopropenium Dimers. *Journal of Chemical Crystallography* **2019**, 1-16.

(49) Polfer, N. C.; Oomens, J. Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass spectrometry reviews* **2009**, *28*, 468-494.

Chapter 6

A Temperature-dependent Blue shifting of OH Stretching Frequencies in Crystalline Cellulose Explained

---

"If one way be better than another, that you may be sure is nature's way."

Aristotle

Part of the material contained within this chapter has previously been published in a peer-reviewed journal article. The citation for this article is stated below.

- Ferreras Moreno, M.; Stinson, C. J.; Shepherd, R. G.; Welsh, I. D.; Altaner, C. M.; Crittenden, D. L. Temperature-dependent Blueshifting of OH Stretching Frequencies in Crystalline Cellulose Explained. *The Journal of Physical Chemistry B* **2020**.

The contributions of each author are the following. Deborah Crittenden was responsible for the computational design, writing the manuscript and supervising the project. Manuel Ferreras Moreno was the chief data processor, responsible for the computational design and running all simulations. Chris Stinson and Ross Shepherd wrote the Fourier transform and Temporal Principal Component Analysis (tPCA) scripts, respectively, to analyse the data. Ivan Welsh ran the tPCA analyses. Clemens Altaner also supervised the project and provided feedback on the manuscript at its draft stage.

The authors thank Pan Chen, Yoshiharu Nishiyama, Jean-Luc Putaux and Karim Mazeau for sharing their GROMACS input files with us.

6.1. Introduction

Cellulose is an abundant and versatile biopolymer.[1] Its fundamental structural unit is a linear polymer of D-glucose, exhibiting intramolecular hydrogen bonding within and between glucose subunits. These chains form into sheets held together by intermolecular hydrogen bonds. These sheets stack to form crystalline fibrils, with relatively weak van der Waals interactions and C–H···O hydrogen bonds between layers.[2,3]

Intra-chain and inter-chain hydrogen bonds between glucose subunits form intricate hydrogen bonding networks that underpin cellulose's ability to form different crystalline allomorphs, which differ according to hydrogen bond configuration.[4] These hydrogen bonding networks also confer a range of desirable physical and chemical properties, including mechanical strength, pliability, hydrophilicity, solubility, thermal stability, sorption capacity and its ability to be derivatized and/or nanostructured.[1,5-7]

Cellulose I is the most common naturally occurring allomorph.[4] The hydrogen bonding patterns of cellulose I have been elucidated by neutron diffraction crystallography, revealing two different hydrogen bonding patterns depending on the organismal source.[2,3] Of these, cellulose Iβ (derived from plants) is the most abundant, widely used and intensively studied. The structure and hydrogen bonding network for a pair of chains within cellulose Iβ is illustrated in Figure 1.[2]

While X-ray and neutron diffraction crystallography are ideal techniques for characterizing static structures,[2,3,8] infrared (IR) spectroscopy can be used to interrogate bonding and dynamics.[9] In particular, it provides an ideal tool for correlating changes in microscopic properties (bond strengths, bonding patterns) with changes in macroscopic variables/conditions (applied strain, temperature, pressure).[10,11] Although it is difficult to definitively assign the IR

spectrum of cellulose due to extensive coupling between vibrational modes,[10,12-14] it is nonetheless straightforward to define spectral regions and monitor changes to band centres and intensities within them.[11,15]

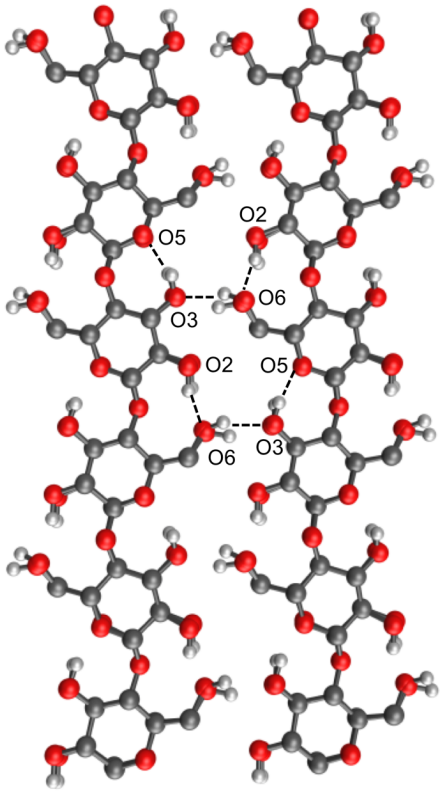


Figure 1: Hydrogen bonding within and between chains in cellulose Iβ from low-temperature neutron crystallography experiments. Subtle differences in hydrogen atom positions due to alternating packing arrangements between sheets can be seen at low temperatures, but these become indistinguishable at higher temperatures. Non-polar hydrogens have been omitted for clarity.

Previous temperature-dependent IR spectroscopy studies of cellulose Iβ[11,16-20] agree that fundamental O-H stretching frequencies increase with temperature, i.e. these bonds get shorter and stronger when cellulose is heated. Initially, this seems counterintuitive because bonds usually weaken and elongate, or even break, upon heating. However, this effect has also been observed for molecules form hydrogen-bonded dimers in solution, where it is convincingly explained by the Finch-Lippincott model.[21] According to this model, O-H bonds strengthen with increasing temperature because the competing H···O hydrogen bonds weaken, and this

effect significantly outweighs any direct thermal weakening of O–H bonds. Mathematically, this is described using a one-dimensional model based upon a potential energy function that contains terms sensitive to the covalent bond length, hydrogen bond length and overall O–H···O distance.[22]

In the context of solution-phase hydrogen-bonded dimers, covalent and hydrogen bonding coordinates are easy to define because the hydrogen bond is linear and the hydrogen bonding coordinate maps directly to an intermolecular stretching mode. However, for hydrogen bonding networks within crystalline cellulose, things get more complicated. It is no longer straightforward – or even necessarily possible – to map hydrogen bonding coordinates to equivalent intra- or inter-molecular stretching modes.

This raises the question – is it still possible to qualitatively rationalise trends in temperature-depending shifting of O–H stretching frequencies in cellulose using the principles that underpin the Finch-Lippincott model? The aim of this work is to answer this question using molecular dynamics simulations.

6.2. Methods

6.2.1. Model Details

Two models have been used in the simulations. The first model (A) was developed by Chen et al[23]. GROMACS input files for this supercell model were obtained from the authors. Briefly, each supercell comprises 6 sheets of 6 chains, each of length 8 residues. At the supercell boundaries, all "dangling bonds" are reattached to periodic neighbouring atoms, according to the minimum image convention. The GROMOS 53a6 force field with the

carbohydrate-specific corrections of Hansen and Hünenburger[24,25] has been supplemented by additional parameters describing aliphatic C–H bonds, required to ensure that experimental geometries and unit cell parameters are correctly reproduced for a range of cellulose allomorphs.[26] Atomic coordinates, topology and force field files are provided in Appendix E.

The second model (B) was developed by ourselves, starting from the crystal structure elucidated by Nishiyama et al.[2] A topology file for the infinitely periodic system was generated with a python-based code that implements the force field parameters extracted from the GROMOS 53a6 force field with the carbohydrate-specific corrections of Hansen and Hünenburger.[24,25] Like Model A, the primitive cell for Model B also comprises 6 sheets of 6 chains each of length 8 residues, but without aliphatic hydrogens. The code, together with the instructions and the topology file, have been included in Appendix E.

6.2.2. Computational Details

One of the primary objectives of this work is to monitor how local-mode vibrational frequencies change as a function of temperature, by monitoring changes in bond lengths as a function of time. This requires well optimized and very well equilibrated simulation conditions, so that sampling simulations can maintain their temperature for a number of vibrational periods even in the absence of an external thermostat. Therefore, we have employed a rigorous energy minimization and thermal equilibration regime.

A 6 step energy minimization process used increasingly fine step sizes with increasingly sophisticated search algorithms to ensure tight (but not complete) convergence to the minimum energy conformation for the model system described above. At each temperature of interest (100 K, 200 K, 300 K, 400 K), the energy-minimized system was gradually brought to the

target temperature by setting the initial atomic velocities to zero and applying the velocity-rescaling thermostat throughout. Pressure control was achieved by simultaneously applying the Berendsen barostat, overall sampling the NPT ensemble. Each equilibration simulation was run for 0.5 ns (500,000 steps with a step size of 1 fs).

Short sampling runs (30 ps, 30,000 frames) were carried out at constant volume and energy, followed by re-equilibration for 50 ps. This process was repeated 20 times, to ensure the results obtained were robust with respect to equilibration conditions. All simulations were performed using the GROMACS program package.[27] Molecular dynamics control files are provided in Appendix E. The same control files were used with each force field.

6.2.3. Analysis

During each sampling calculation, positions of all atoms in the primitive supercell were recorded at every time step. From these geometries, 20 unique O2–H2, O3–H3 and O6–H6 bond lengths and corresponding O···O hydrogen bond distances and O–H···O hydrogen bond angles were extracted from both models, and their average values computed. Local-mode vibrational frequencies were extracted by selecting the highest amplitude peak from the discrete Fourier transform of the O–H bond length data.

Similarly, 20 unique chain-twist dihedrals, φ, were identified as illustrated in Figure 2, and their mean signed and unsigned values computed and stored. Each simulation yielded a single set of mean average values for the predicted crystallographic lattice parameters at the specified temperature and atmospheric pressure.

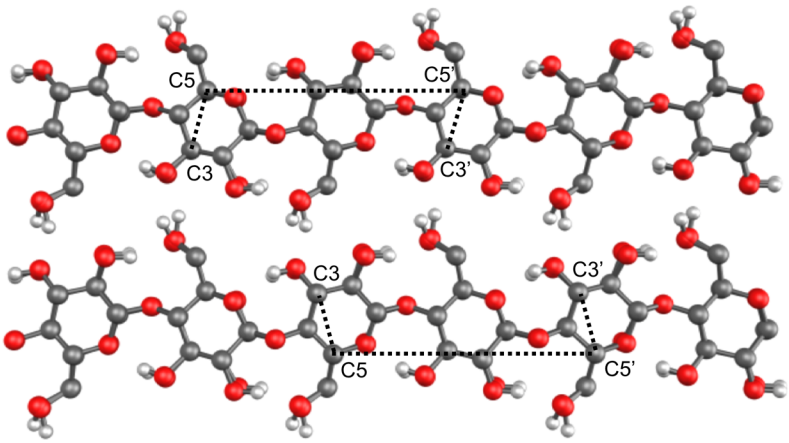


Figure 2: Definition of chain-twist coordinate.[28]

To ensure statistical significance, 20 independent sampling simulations were carried out at each temperature, yielding a total of 400 observations of O–H stretching frequencies, time-averaged O–H distances, O–H· · ·O angles, O···O distances and C3–C5–C5'–C3' chain-twist dihedral angles, and 20 independent sets of lattice parameters (a = inter- sheet direction, b = inter-chain direction, c = along-chain direction and γ = trapezoidal distortion parameter). Finally, averages of these values were computed, along with the standard error of the mean:

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{N\,(N-1)} \sum_{i=1}^{N} (x_i - \bar{x})^2} \qquad (6.1)$$

where N = the number of independent observations, x_i is the i^{th} value of the variable of interest, and $\bar{x}$ its mean value.

Temperature-dependent changes in final averaged values for both models were quantified using linear regression where appropriate. All values are reported as slope ± standard error of the slope.

6.3. Results and Discussion

6.3.1. Local Vibrational and Geometric Parameters

The temperature dependence of O–H stretching frequencies, along with vibrationally-averaged O–H bond lengths, O···O distances and O–H···O angles for models A and B are illustrated in Figure 4A & 4B and summarized in Table 1A & 1B, respectively.

Table 1A: Temperature-dependence of O−H stretching frequencies, O−H and O···O distances, and O-H-O angles in crystalline cellulose Iβ (Model A).

Bonds	$\partial_T\{\upsilon_{\text{O-H}}\}$ ($\text{cm}^{-1}\ \text{K}^{-1}$)	$\partial_T\{r_{\text{O-H}}\}$ ($\times10^{-6}$ Å K^{-1})	$\partial_T\{r_{\text{O}\cdots\text{O}}\}$ ($\times10^{-6}$ Å K^{-1})	$\partial_T\{\theta_{\text{OHO}}\}$ ($\times10^{-3}$ ° K^{-1})
O6–H6···O3	0.018 ± 0.004	–5.1 ± 0.1	92.8 ± 1.0	–8.3 ± 0.1
O2–H2···O6	0.013 ± 0.002	–1.8 ± 0.7	73.9 ± 3.2	–3.8 ± 0.1
O3–H3···O5	0.010 ± 0.003	–3.3 ± 0.1	68.2 ± 0.5	–7.0 ± 0.1

Table 1B: Temperature-dependence of O−H stretching frequencies, O−H and O···O distances, and O-H-O angles in crystalline cellulose Iβ (Model B).

Bonds	$\partial_T\{\upsilon_{\text{O-H}}\}$ ($\text{cm}^{-1}\ \text{K}^{-1}$)	$\partial_T\{r_{\text{O-H}}\}$ ($\times10^{-6}$ Å K^{-1})	$\partial_T\{r_{\text{O}\cdots\text{O}}\}$ ($\times10^{-6}$ Å K^{-1})	$\partial_T\{\theta_{\text{OHO}}\}$ ($\times10^{-3}$ ° K^{-1})
O6–H6···O3	0.008 ± 0.001	–2.2 ± 0.1	75.0 ± 1.0	–6.5 ± 0.1
O2–H2···O6	0.009 ± 0.001	–4.7 ± 0.1	80.2 ± 1.7	–3.6 ± 0.1
O3–H3···O5	0.019 ± 0.003	–4.6 ± 0.1	48.2 ± 1.5	–6.3 ± 0.1

Consistent with temperature-dependent infrared spectroscopy studies of cellulose,[11] covalent O–H stretching frequencies increase with temperature (Figures 4A-a & 4B-a), although the size of the predicted change in both models (∼0.01 $\text{cm}^{-1}\ \text{K}^{-1}$) is an order of magnitude smaller than that observed experimentally (∼0.1 $\text{cm}^{-1}\ \text{K}^{-1}$). However, it is well known that bonded terms within molecular mechanics force fields tend to underestimate anharmonicity and overestimate stiffness of covalent bonds,[29] hence reducing sensitivity to thermally-induced shifts. In addition, we are using "infinite crystal" models with strict periodic boundary

conditions that also limit the flexibility of the system. Despite these limitations in the models, we nonetheless achieve qualitative agreement with experiment, which allows us to use them to investigate how thermally-induced structural changes within cellulose lead to the observed weakening of O–H bonds with temperature. Small differences have been observed in the predictions between the two models employed. For the intermolecularly hydrogen-bonded O6–H6···O3 covalent stretching fundamental, Model A predicts a larger thermal shift than model B, by about a factor of 2. For the intramolecularly hydrogen bonded O3–H3···O5 covalent stretching frequency, the opposite trend is observed.

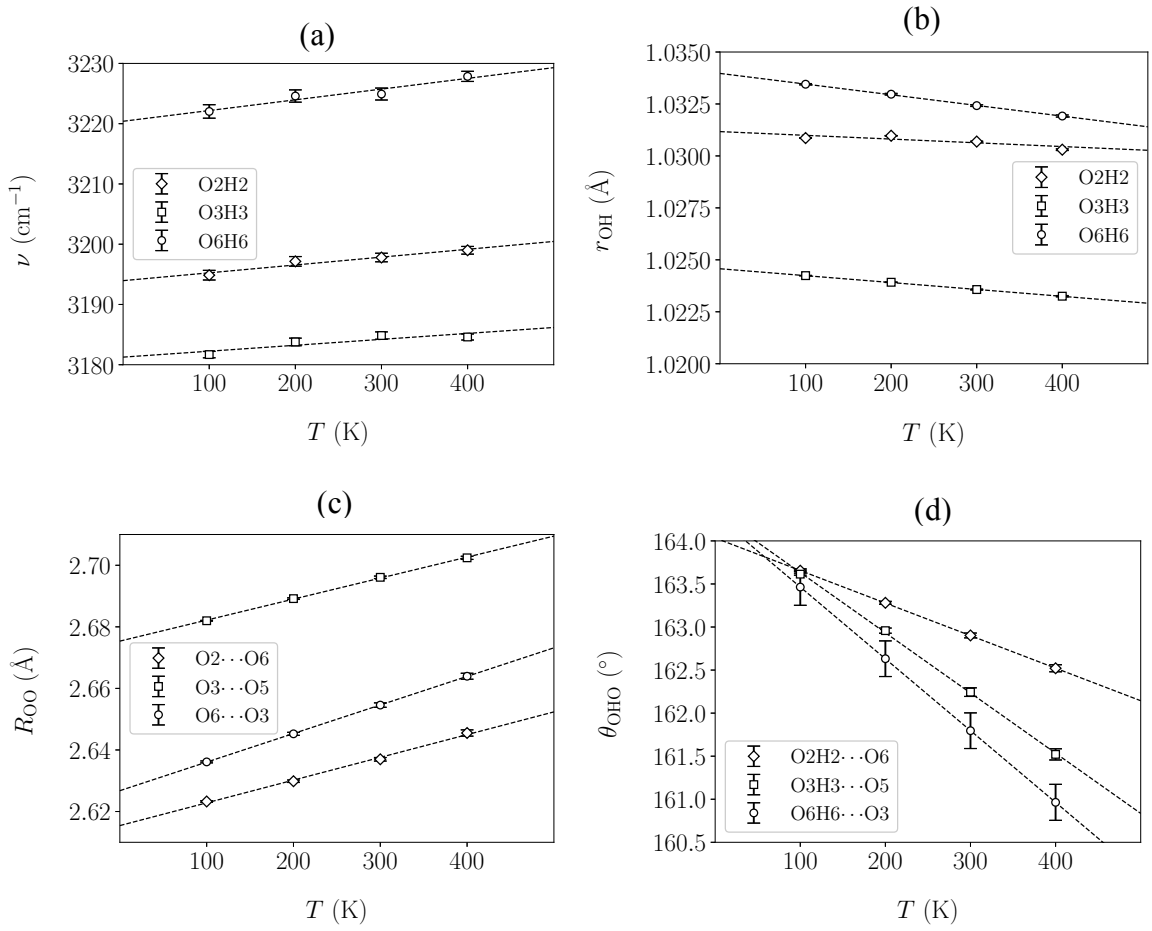


Figure 4A: Temperature dependence of (a) O–H stretching frequencies and associated local geometric parameters, (b) O–H bond lengths, (c) O···O distances and (d) O–H–O angles for model A.

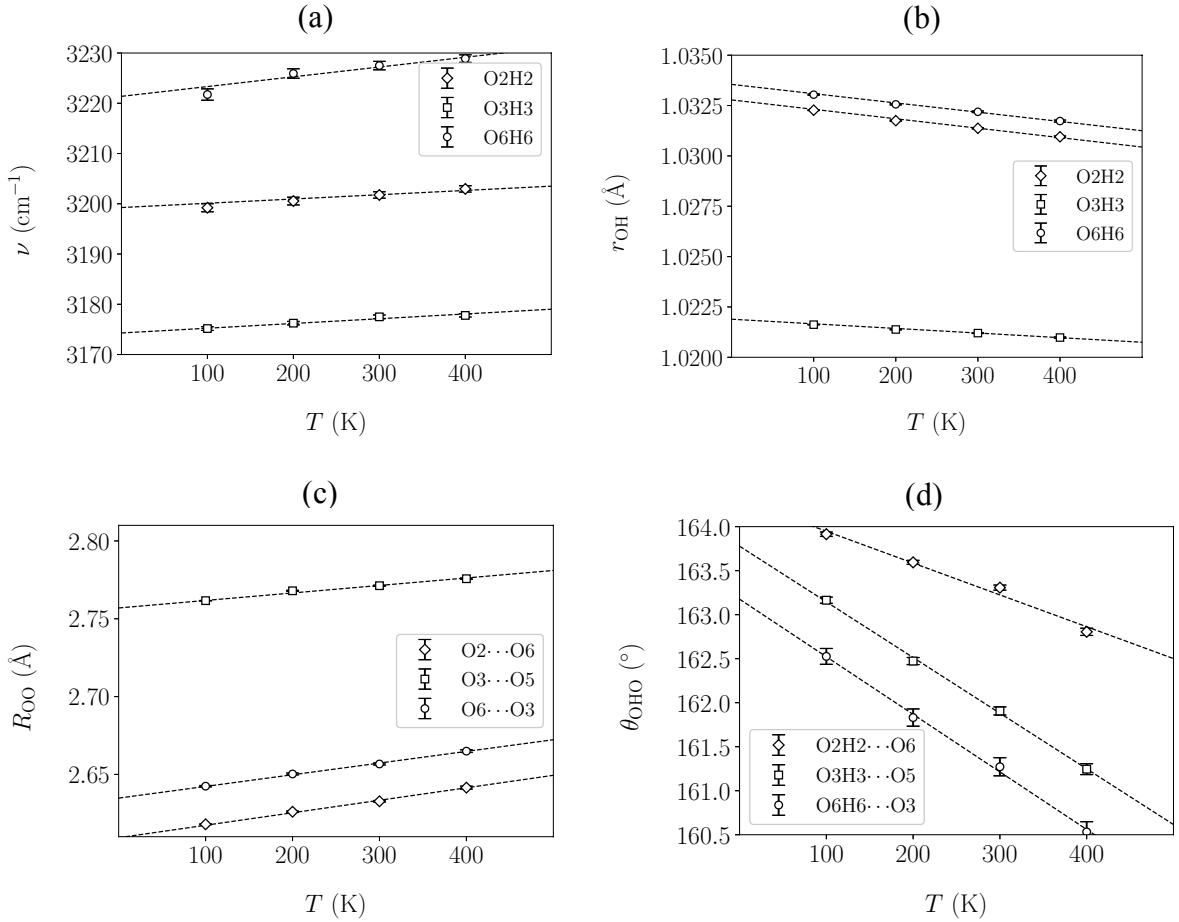


Figure 4B: Temperature dependence of (a) O–H stretching frequencies and associated local geometric parameters, (b) O–H bond lengths, (c) O···O distances and (d) O–H–O angles for model B.

Consistent with the Finch-Lippincott model, covalent O–H bonds get stronger (Figures 4A-a & 4B-a) and shorter (Figures 4A-b & 4B-b) with increasing temperature due to weakening of the associated hydrogen bonds. This can be inferred from the geometric data presented in Figures c and d for both models. As the temperature increases, O···O distances get longer and O–H···O angles get more acute. Both of these effects contribute to making non-bonded H···O distances longer and non-bonded electrostatic interactions weaker.

While this is a useful description of what changes occur and what they mean for hydrogen bond strengths, it does not explain how these geometric changes arise when cellulose is heated. One thing is clear – the mechanism is not the same as for linear hydrogen bonds formed between isolated molecules in solution, in which the intermolecular stretching mode is also the hydrogen

bonding coordinate. Thermal excitation of the intermolecular stretching mode leads to longer O···O distances, weakening the hydrogen bond and strengthening the covalent bond. However, in cellulose, there is a complicated hydrogen bonding network and it is not possible to define simple one-dimensional hydrogen bond stretching coordinates, not least because the hydrogen bonds are bent (Figures 4A-d & 4B-d).

Instead, possible explanations include: thermally-induced chain separation (would stretch and weaken inter-chain hydrogen bonds), thermally-induced chain stretching (would elongate and weaken intra-chain hydrogen bonds), increased chain twisting at higher temperatures (would affect entire hydrogen bonding network), or local changes to the conformation of the glucose ring and/or orientation of glucose units relative to one another (would affect intra-chain hydrogen bonds but the direction of the change is not obvious *a priori*).

These considerations prompt the following questions: is elongation of the cellulose chain along crystallographic *c* axis related to the observed lengthening of O6–O2 and O3–O5 distances? Is chain separation responsible for/related to the observed lengthening of O6–O3 bonds? are changes in sheet separation correlated with changes in hydrogen bond strengths and O···O lengths? In order to find insight into what is causing the shifts in local O–H stretching frequencies and O–H/O–O bond lengths, we will examine how global geometric parameters change with temperature.

6.3.2. Global Geometric Parameters

Temperature-dependent changes in crystallographic lattice parameters and chain-twist angles for both models are illustrated in Figure 5A & 5B and summarized in Table 2A & 2B, respectively

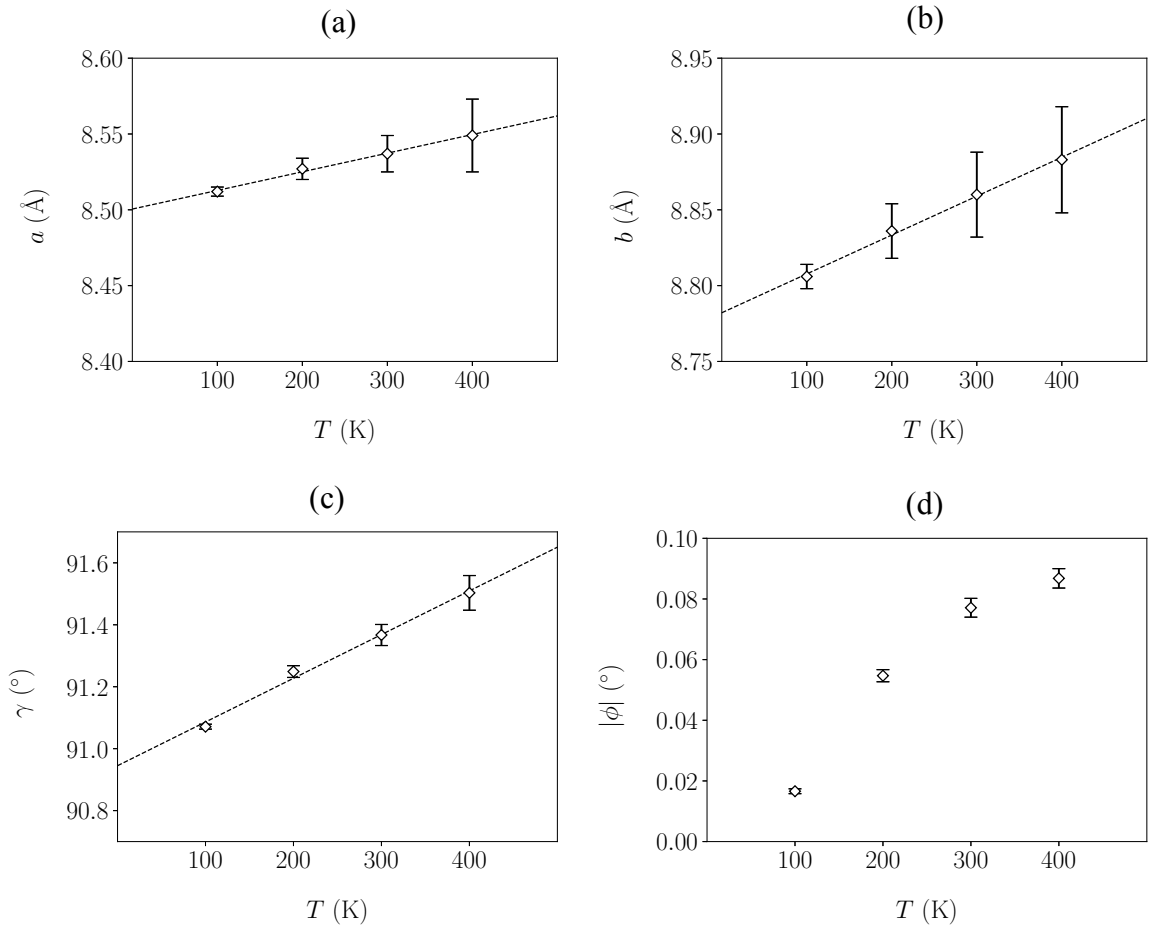
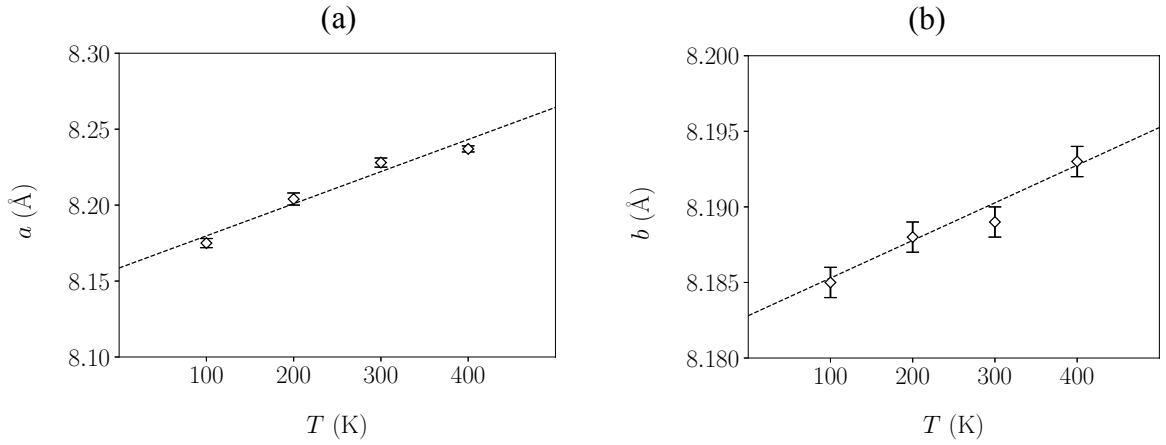


Figure 5A: Temperature dependence of global geometric parameters: crystallographic lattice parameters (a-c) and mean absolute chain-twist angles (d). No significant variation in the along-chain crystallographic lattice parameter c was observed so only results for (a) inter-sheet spacings, (b) inter-chain spacings and (c) trapezoidal distortion are reported here.



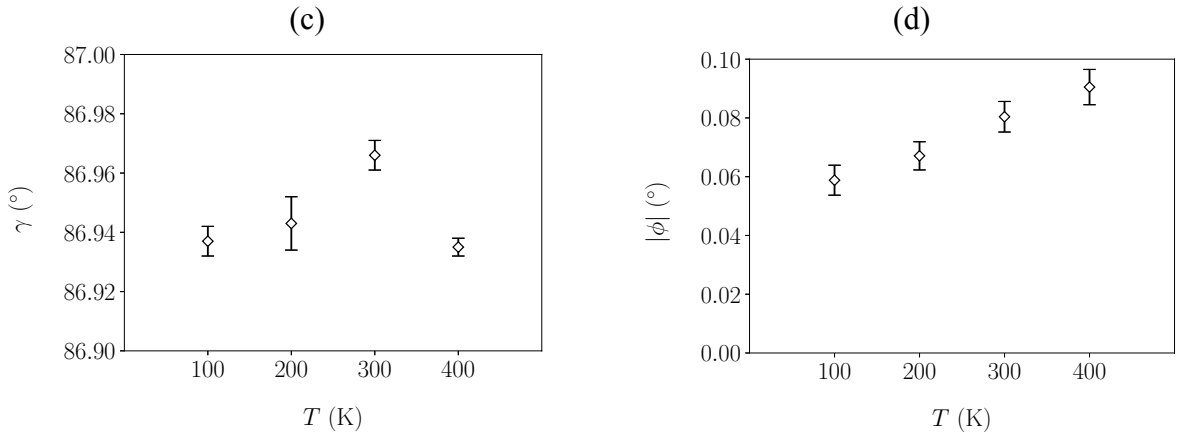


Figure 5B: Temperature dependence of global geometric parameters: crystallographic lattice parameters (a-c) and mean absolute chain-twist angles (d). No significant variation in the along-chain crystallographic lattice parameter c was observed so only results for (a) inter-sheet spacings, (b) inter-chain spacings and (c) trapezoidal distortion are reported here.

No detectable changes in expansion/contraction along cellulose chains in either model (along cellulose chains) were observed to within the statistical resolution of our simulations. This is consistent with the results of previous variable-temperature X-ray diffraction studies,[30] which show that expansion along the *c* direction with temperature is 1–2 orders of magnitude smaller than along the *a* (inter-sheet) and *b* (inter-chain) directions. It is also consistent with previous infrared spectroscopy studies which infer from mechanical stretching experiments that chain elongation cannot fully (or even mostly) explain the observed blue-shift in intramolecularly hydrogen-bonded O–H stretching frequencies.[11]

Table 2A: Temperature dependent shifts in crystalline lattice parameters for crystalline cellulose Iβ (Model A).

Thermal gradient	**This work**	**Computational**			**Experimental**	
		Ref[13]	Ref[31]	Ref[32]	Ref[33]	Ref[34]
$\partial_T\{a\}$ (x10^{-6} Å K^{-1})	123 ± 7	108	73	830	98	52
$\partial_T\{b\}$ (x10^{-6} Å K^{-1})	256 ± 12	13	68	250	12	21
$\partial_T\{c\}$ (x10^{-6} Å K^{-1})	~ 0	–	–	~ 0	–	4
$\partial_T\{\gamma\}$ (x10^{-3} ° K^{-1})	1.4 ± 0.1	–	–	-0.625	-0.625	-0.25

Table 2B: Temperature dependent shifts in crystalline lattice parameters for crystalline cellulose Iβ (Model B).

Thermal gradient	This work	Computational			Experimental	
		Ref[13]	Ref[31]	Ref[32]	Ref[33]	Ref[34]
$\partial_T\{a\}$ (x10^{-6} Å K^{-1})	211 ± 3	108	73	830	98	52
$\partial_T\{b\}$ (x10^{-6} Å K^{-1})	24.9 ± 4.1	13	68	250	12	21
$\partial_T\{c\}$ (x10^{-6} Å K^{-1})	~ 0	–	–	~ 0	–	4
$\partial_T\{\gamma\}$ (x10^{-3} ° K^{-1})	~ 0	–	–	-0.625	-0.625	-0.25

Similarly, the average chain-twist angle in both models remained at zero regardless of temperature, as expected due to the infinitely periodic nature of the system. The magnitude of the mean absolute chain-twist angles increases only marginally with temperature, to a maximum of just under 0.1° in both models before the onset of thermal decomposition around 500 K.[35] It is unlikely that these small, transient, random-direction chain-twisting motions within crystalline cellulose have any substantive effect on disrupting its hydrogen bonding network. Therefore, chain twisting also does not explain the observed changes in O···O distances and O–H···O angles.

Both models underestimate the unit cell trapezoidal distortion (γ) by $\sim 7^{o}$ in model A and $\sim 10^{o}$ in model B, as we can see in Table 3. Temperature-induced changes are small according to model A and non-existent according to model B. These results are also consistent with previous experimental observations. Therefore, rearrangement of the cellulose fibrils relative to one another also does not explain why hydrogen bonds weaken at higher temperatures.

Table 3: Comparison of global geometric parameters at 295K. Values for Model A & B were obtained by interpolation.

Parameters	Model A	Model B	Experimental[2]
a (Å)	8.53	8.22	7.76
b (Å)	8.87	8.19	8.20
c (Å)	10.43	7.86	10.37
γ (deg)	89.47	86.96	96.62

Finally, we look at the two key parameters that do change appreciably with temperature: the sheet separation unit cell parameter (*a*) and the chain separation unit cell parameter (*b*). Predictions from Model B are in better agreement with experiment than those of Model A (Table 3). Model B also predicts thermal shifts of the correct order of magnitude along both axes, whereas model A overestimates the change in *b* by an order of magnitude (Table 2A & 2B). From this results we can infer that model B produces a more realistically shaped potential energy surface, possibly due to the fact that model A has been optimized to get the unit cell parameters correct – adding the aliphatic hydrogens improves how well the model recovers the sheet separation parameter (a).[26] This improvement in geometric accuracy apparently results in a decrease in energetic accuracy.

Regardless of the differences in the values of the thermal shifts between models, both of our predictions are qualitatively consistent with previous experimental and computational observations, i.e. they both show increases in chain and sheet spacing with increasing temperature. The thermally-induced chain separation along the *b* axis explains the observed increase in O6···O3 distances with temperature, weakening the H6···O3 hydrogen bond and strengthening the O6–H6 covalent bond. Increases in O2···O6 and O3···O5 distances with temperature are harder to understand, but it is plausible that they could be connected to the observed thermal expansion in the *a* direction, i.e. thermally-induced separation of the cellulose sheets. For example, sheet separation could be driven by excitation of vibrational modes that break the overall planarity of the sheets, e.g. rotation about the C5–C6 bond of the hydroxymethyl ring substituents, enhanced ring puckering, or reorientation of the glucose units relative to one another via twisting or hinging type motions of the connecting glycosidic bridge.

6.4. Conclusions

Using molecular dynamics simulations, we have qualitatively reproduced the observed blue-shift in covalent O–H stretching frequencies within infinitely repeating crystalline cellulose, and shown that this is correlated to weakening of associated hydrogen bonds, which become longer and less linear. For inter-chain hydrogen bonds, the mechanism behind this is relatively straightforward. Direct thermal excitation of inter-chain separation modes leads to increased chain spacing and therefore weaker hydrogen bonds. However, for intra-chain hydrogen bonds, the mechanism is less immediately apparent. The only thing that is immediately clear is that weakening of hydrogen bonds along the chain is *not* due to thermal expansion along the length of the chain. Instead, it is more likely that hydrogen bond lengthening and weakening is caused by increased ring puckering/distortion at higher temperatures.

References

(1) Qiu, X.; Hu, S. “Smart” materials based on cellulose: a review of the preparations, properties, and applications. *Materials* **2013**, *6*, 738-781.

(2) Nishiyama, Y.; Langan, P.; Chanzy, H. Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction. *Journal of the American Chemical Society* **2002**, *124*, 9074-9082.

(3) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. Crystal structure and hydrogen bonding system in cellulose Iα from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society* **2003**, *125*, 14300-14306.

(4) O'sullivan, A. C. Cellulose: the structure slowly unravels. *Cellulose* **1997**, *4*, 173-207.

(5) Shen, T.; Gnanakaran, S. The stability of cellulose: a statistical perspective from a coarse-grained model of hydrogen-bond networks. *Biophysical journal* **2009**, *96*, 3032-3040.

(6) Chundawat, S. P.; Bellesia, G.; Uppugundla, N.; da Costa Sousa, L.; Gao, D.; Cheh, A. M.; Agarwal, U. P.; Bianchetti, C. M.; Phillips Jr, G. N.; Langan, P. Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. *Journal of the American Chemical Society* **2011**, *133*, 11163-11174.

(7) Tayeb, A. H.; Amini, E.; Ghasemi, S.; Tajvidi, M. Cellulose nanomaterials—Binding properties and applications: A review. *Molecules* **2018**, *23*, 2684.

(8) Langan, P.; Sukumar, N.; Nishiyama, Y.; Chanzy, H. Synchrotron X-ray structures of cellulose I β and regenerated cellulose II at ambient temperature and 100 K. *Cellulose* **2005**, *12*, 551-562.

(9) Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. Watching hydrogen-bond dynamics in a β-turn by transient two-dimensional infrared spectroscopy. *Nature* **2006**, *444*, 469-472.

(10) Altaner, C. M.; Thomas, L. H.; Fernandes, A. N.; Jarvis, M. C. How cellulose stretches: synergism between covalent and hydrogen bonding. *Biomacromolecules* **2014**, *15*, 791-798.

(11) Altaner, C. M.; Horikawa, Y.; Sugiyama, J.; Jarvis, M. C. Cellulose Iβ investigated by IR-spectroscopy at low temperatures. *Cellulose* **2014**, *21*, 3171-3179.

(12) Maréchal, Y.; Chanzy, H. The hydrogen bond network in Iβ cellulose as observed by infrared spectrometry. *Journal of Molecular Structure* **2000**, *523*, 183-196.

(13) Agarwal, V.; Huber, G. W.; Jr., W. C. C.; Auerbach, S. M. Simulating infrared spectra and hydrogen bonding in cellulose Iβ at elevated temperatures. *The Journal of Chemical Physics* **2011**, *135*, 134506.

(14) Lee, C. M.; Kubicki, J. D.; Fan, B.; Zhong, L.; Jarvis, M. C.; Kim, S. H. Hydrogen-Bonding Network and OH Stretch Vibration of Cellulose: Comparison of Computational Modeling with Polarized IR and SFG Spectra. *The Journal of Physical Chemistry B* **2015**, *119*, 15138-15149.

(15) Makarem, M.; Lee, C. M.; Kafle, K.; Huang, S.; Chae, I.; Yang, H.; Kubicki, J. D.; Kim, S. H. Probing cellulose structures with vibrational spectroscopy. *Cellulose* **2019**, *26*, 35-79.

(16) Katon, J.; Miller Jr, J.; Bentley, F. The infrared spectra of carbohydrates at sub-ambient temperatures. *Carbohydrate Research* **1969**, *10*, 505-516.

(17) Michell, A. J. Low-temperature infrared spectra of carbohydrates. *Australian Journal of Chemistry* **1970**, *23*, 833-838.

(18) McCall, E. R.; Morris, N. M.; Tripp, V. W.; O'Connor, R. T. Low temperature infrared absorption spectra of cellulosics. *Applied Spectroscopy* **1971**, *25*, 196-200.

(19) Hatakeyama, H.; Nagasaki, C.; Yurugi, T. Relation of certain infrared bands to conformational changes of cellulose and cellulose oligosaccharides. *Carbohydrate research* **1976**, *48*, 149-158.

(20) Ivanova, N.; Korolik, E.; Zhbankov, R.; Insarova, N. Low-temperature IR spectra of cellulose. *Journal of Applied Spectroscopy* **1983**, *38*, 676-678.

(21) Finch, J. N.; Lippincott, E. R. Hydrogen Bond Systems. Temperature Dependence of OH Frequency Shifts and OH Band Intensities. *The Journal of Physical Chemistry* **1957**, *61*, 894-902.

(22) Lippincott, E. R.; Schroeder, R. One-dimensional model of the hydrogen bond. *The Journal of Chemical Physics* **1955**, *23*, 1099-1106.

(23) Chen, P.; Nishiyama, Y.; Putaux, J.-L.; Mazeau, K. Diversity of potential hydrogen bonds in cellulose I revealed by molecular dynamics simulation. *Cellulose* **2014**, *21*, 897-908.

(24) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry* **2004**, *25*, 1656-1676.

(25) Hansen, H. S.; Hünenberger, P. H. A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. *Journal of computational chemistry* **2011**, *32*, 998-1032.

(26) Chen, P.; Nishiyama, Y.; Mazeau, K. Atomic partial charges and one Lennard-Jones parameter crucial to model cellulose allomorphs. *Cellulose* **2014**, *21*, 2207-2217.

(27) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **2008**, *4*, 435-447.

(28) Bu, L.; Himmel, M. E.; Crowley, M. F. The molecular origins of twist in cellulose I-beta. *Carbohydrate polymers* **2015**, *125*, 146-152.

(29) Shannon, R.; Hornung, B.; Tew, D.; Glowacki, D. Anharmonic Molecular Mechanics: Ab Initio Based Morse Parametrizations for the Popular MM3 Force Field. *The Journal of Physical Chemistry A* **2019**, *123*, 2991-2999.

(30) Diaz, J. A.; Wu, X.; Martini, A.; Youngblood, J. P.; Moon, R. J. Thermal expansion of self-organized and shear-oriented cellulose nanocrystal films. *Biomacromolecules* **2013**, *14*, 2900-2908.

(31) Bergenstråhle, M.; Berglund, L. A.; Mazeau, K. Thermal response in crystalline Iβ cellulose: a molecular dynamics study. *The Journal of Physical Chemistry B* **2007**, *111*, 9138-9145.

(32) Zhang, Q.; Bulone, V.; Ågren, H.; Tu, Y. A molecular dynamics study of the thermal response of crystalline cellulose Iβ. *Cellulose* **2011**, *18*, 207-221.

(33) Wada, M.; Hori, R.; Kim, U.-J.; Sasaki, S. X-ray diffraction study on the thermal expansion behavior of cellulose Iβ and its high-temperature phase. *Polymer Degradation and Stability* **2010**, *95*, 1330-1334.

(34) Hidaka, H.; Kim, U.-J.; Wada, M. Synchrotron X-ray fiber diffraction study on the thermal expansion behavior of cellulose crystals in tension wood of Japanese poplar in the low-temperature region. *Holzforschung* **2010**, *64*, 167-171.

(35) Yang, H.; Yan, R.; Chen, H.; Lee, D. H.; Zheng, C. Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel* **2007**, *86*, 1781-1788.

Chapter 7

Conclusions and Future Work

"The end of a melody is not its goal: but nonetheless, had the melody not reached its end it would have not reached its goal either. A parable."

Friedrich Wilhelm Nietzsche

7.1 Conclusions

The aim of this thesis was to determine how effectively computational chemistry tools can be applied to assign, interpret and explain vibrational spectra of molecules and materials.

In Chapter 3, we showed that fundamental frequencies from gas phase IR spectra of small semi-rigid molecules can be quantitatively assigned, provided that force fields are constructed using high-level electronic structure models and the nuclear vibrational Schrödinger equation solved by performing fully converged vibrational configuration interaction calculations. However, constructing force fields at high levels of electronic structure theory is not always computationally feasible, so we also investigated a "hybrid" approach in which anharmonic force fields were computed at a relatively low level of theory (HF/cc-pVTZ and MP/cc-pVTZ) and their harmonic force constants replaced with values obtained at a much higher level of theory (CCSD(T)/cc-pVTZ or higher). Our results indicate that the commonly-used approach of constructing hybrid force fields in which the harmonic part of the potential energy surface is computed at a high level of *ab initio* theory and the anharmonic counterpart at a lower level is well justified. In particular, we recommend using MP2/cc-pVTZ to compute anharmonic force fields, and then replacing the harmonic force constants with values obtained at the CCSD(T)/cc-pVTZ level of theory or higher. This approach yields fundamental frequencies within 10 cm^{-1} of benchmark values.

Although the hybrid force field approach can extend the range of applicability of anharmonic vibrational frequency calculations, this approach is still limited to relatively small and rigid molecules (< 10 atoms). For larger systems, harmonic vibrational frequency calculations remain the only computationally affordable option. Although it is well known that harmonic nuclear vibrational models cannot achieve quantitative accuracy, they can be useful in

qualitatively assigning spectra, especially for systems in which symmetry plays an important role in determining the nature and intensity of the fundamental IR transitions.

In Chapters 4 & 5, quantum chemical models were used to interpret and assign the experimentally measured IR spectra of two different chloride hydrate clusters trapped within cyclopropenium crystal environments. In Chapter 4, a dichloride tetrahydrate was crystallographically observed trapped as a discrete unit surrounded by bis(diphenylamino) diethylamino cyclopropenium cations. To confirm its discrete nature, we performed gas phase geometry optimizations to confirm that it remained a minimum energy structure even in the absence of any specific stabilizing interactions with the surrounding crystalline environment. To assign the IR spectrum, we computed harmonic fundamental frequencies at MP2/aug-cc-pVDZ and then applied an empirical scaling model to bring them into closer agreement with experiment. The two highest intensity transitions were assigned as coupled symmetric stretches of the "inner" and "outer" water molecules within the cluster. A single lower intensity transition at a higher frequency was assigned to a fully coupled asymmetric stretching mode, involving the concerted displacement of all hydrogen atoms within the system towards the terminal chloride ions. Bending fundamentals were predicted to within 30 cm^{-1} of experiment, and corresponding overtones to within 60 cm^{-1}. As for the symmetric stretching modes, these correspond to coupled bending motions of the "inner" and "outer" water molecules.

In Chapter 5, a series of dichloride dihydrates were crystallographically characterised surrounded by triaminocyclopropenium cations with different alkyl and fluoroalkyl substituents attached. Although the encapsulated dichloride dihydrates were all structurally similar, their IR spectra were quite different depending on the substitution pattern. Therefore, a different approach to simulating the IR spectra was required, in order to capture environmental effects. A model system comprising di- and tri-fluoromethane molecules was

extracted from the crystal structure of the least symmetric complex. An iterative (and alternate) geometry optimization process was performed to optimize the positions of both the fluorocarbons and the dichloride dihydrate cluster, keeping the positions of the carbon atoms fixed throughout. Upon convergence, a partial Hessian calculation was carried out on the dichloride dihydrate. As in the previous study, the same empirical quadratic correction was applied to obtain the predicted anharmonic vibrational frequencies. In this instance, the truncated nature of the model system accentuated the disagreement between predicted and observed transition frequencies, but the spectra could nonetheless still be unambiguously assigned. The two highest frequency stretching bands were assigned to in- and out-of-phase asymmetric stretching modes of the water molecules while a single lower-frequency band was assigned to coincident symmetric stretching modes based on the small difference found in the calculated frequencies and the intensity of the band. Only one of the two expected bending modes was observed experimentally and assigned as an out-of-phase coupled bending mode. For more symmetric complexes, only a subset of these vibrational modes were IR active.

While quantum mechanical calculations are required to predict and assign IR spectra, thermal shifts in stretching frequencies can be simulated classically. This is particularly useful in simulating vibrational motion of large systems for which performing accurate quantum chemical calculations is even more computationally challenging. In the final results chapter, molecular dynamics simulations were performed to investigate why O-H stretching frequencies blue-shift (increase in apparent bond strength) with increasing temperature. Although isolated covalent bonds generally weaken with temperature, in this case, this is more than offset by adjacent hydrogen bonds weakening, leading to an overall strengthening of the the covalent bonds themselves. The mechanism behind this effect varies for inter- and intra-chain hydrogen bonds. Thermally-induced expansion leads to increased separation between the cellulose chains and consequently longer inter-chain hydrogen bond distances and weaker

inter-chain hydrogen bonds. On the other hand, the dihedral principal component analysis reveals that local changes at each glucose unit are behind the weakening of the intra-chain hydrogen bond: rotation of methoxy groups, enhanced puckering of glucose rings, and hinging of bonds directly connected to the pyranose ring system away from their usual equatorial configuration.

Overall, IR spectroscopy provides a very sensitive probe of bonding within and between molecules and materials. For small molecules in the gas phase, fully anharmonic vibrational frequency calculations can be used to exactly simulate IR spectra, conclusively confirming identity of the molecule, its conformation and its bonding structure. For larger molecules and clusters, more approximate models must be employed. Nonetheless, experimental spectra can be qualitatively assigned with a high degree of confidence, particularly if the system under investigation possesses a high degree of symmetry. Finally, thermal effects on observed stretching frequencies can be simulated using classical molecular dynamics models, without needing to fully assign the IR spectrum at all, but still providing substantial physical insight into structure and bonding.

7.2 Future work

Throughout this thesis, we have judiciously applied available computational methods to problems that needed solving. To be able to solve problems that are not currently feasible, new computational tools or approaches would need to be developed. For example, the IR spectrum of cellulose has not yet been convincingly assigned in full or even in part. Cellulose can exist in – and occasionally interconvert between – many allomorphs with different hydrogen bonding patterns. These hydrogen bonding patterns should be reflected in the OH stretching region of its "fingerprint" IR spectrum. It would be useful to be able to simulate the

IR spectrum of cellulose directly from its proposed structure in order to confirm hydrogen bonding patterns. To achieve this, it would be necessary to perform anharmonic vibrational frequency calculations on cellulose. This could, in principle, be achieved by interfacing force field models to describe the potential energy surface with anharmonic quantum nuclear vibrational models like vibrational perturbation theory or vibrational configuration interaction theory. Further, prediction accuracy could be improved by ensuring that harmonic force constants are particularly accurately parameterised.

Another line of research presented in this thesis was the study of chloride hydrate clusters. Simulating the anharmonic vibrational frequencies of these clusters was only achieved with the use of an empirical correction model. Attempting the quantitative assignment of the spectrum of these systems seems a reachable goal, but would involve extending the PyPES program to include intermolecular coordinates for expanding the potential energy surface.

Appendix A

Fundamentals of Infrared and Raman Spectroscopy

Due to the quantum nature of matter, molecules possess discrete rotational and vibrational energy levels that can be measured spectroscopically and/or modelled computationally by solving the rovibrational nuclear Schrödinger equation. However, because rotational and vibrational motions typically occur on substantially different energy and length scales, they are often measured and modelled separately. This thesis particularly focusses on modelling transitions amongst vibrational energy levels, which can be probed spectroscopically using either infrared (IR) absorption or Raman "scattering" (absorption + emission) techniques.

In IR spectroscopy, transitions between vibrational energy levels are probed by directly measuring the energy of the photons absorbed by a sample. A molecule can only absorb photons whose energy is equal to the difference between two vibrational states. This process is indicated by the left-most arrow in Figure 1.

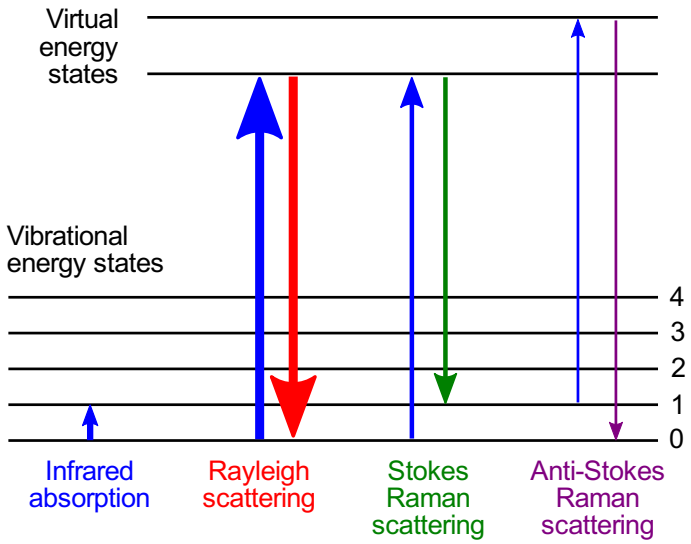


Figure 1. Energy diagram in which the horizontal lines represent different vibrational states and in which the transitions between different energetic states are shown for different light-matter interactions. https://commons.wikimedia.org/w/index.php?curid=7845122

Alternatively, Raman spectroscopy probes transitions between vibrational energy levels indirectly. These transitions are computed from the difference in energy between the incident and "inelastically scattered" photons. When the energy of the photons of the incident beam, $h\nu_o$ (where h is the Planck's constant), is much greater than the difference in energy between two vibrational (or rotational) levels of the molecule, most of photons pass through but a small fraction are absorbed and re-emitted. If the re-emitted photon is at the same wavelength as the absorbed photon, this process is referred to as Rayleigh scattering, and no information can be gained about the vibrational states of the molecule. This process is illustrated by the blue and red arrows in Figure 1.

However, if the molecule undergoes a transition to a different vibrational state than it was in initially, then information about the vibrational energy level spacings can be gained from the difference in frequencies of the absorbed and emitted photons. The most common situation is Stokes scattering, in which the molecule is excited from its vibrational ground state and transitions to an excited state, emitting a lower frequency photon in the process (blue and green arrows in Figure 1), primarily because most molecules exist in their ground vibrational state under common experimental conditions (at or below room temperature). However, at higher temperatures, a substantial fraction of molecules may start in a vibrationally-excited state; in this case anti-Stokes scattering may occur (blue and purple arrows in Figure 1).

Appendix B

Benchmarking Potential Energy Surface Construction Scheme Data

Contents

15. **Table 15.** Reported fundamental vibrational frequencies from original publications and computed values from this work of NH_4^+.
16. **Table 16.** Reported fundamental vibrational frequencies from original publications and computed values from this work of PH_3.
17. **Table 17.** Reported fundamental vibrational frequencies from original publications and computed values from this work of SiF_2.
18. **Table 18.** Reported fundamental vibrational frequencies from original publications and computed values from this work of SiH_4.
19. **Table 19.** Reported fundamental vibrational frequencies from original publications and computed values from this work of SO_2.
20. **Table 20.** Reported fundamental vibrational frequencies from original publications and computed values from this work of SO_3.

References

Table 1. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of C_2H_4.

Reference[1]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
824.8	892.30	834.78	829.78	828.12	826.20	826.13
936.4	1071.17	956.60	940.25	933.03	935.71	936.42
951.8	1088.41	977.67	962.00	945.54	948.76	948.94
1027.0	1122.39	1061.04	1030.66	1026.77	1030.69	1030.02
1225.8	1323.93	1233.59	1229.26	1225.33	1231.13	1231.42
1343.9	1452.64	1367.31	1352.29	1340.46	1341.91	1342.32
1442.8	1556.10	1457.02	1451.38	1438.25	1440.95	1441.45
1626.8	1776.36	1650.40	1637.92	1620.91	1625.44	1626.17
2987.9	3145.19	3045.85	3008.54	2978.12	2986.53	2988.09
3017.8	3178.77	3089.69	3048.88	3010.98	3018.86	3018.80
3081.9	3229.23	3139.13	3085.07	3073.98	3088.97	3088.68
3103.8	3245.69	3165.40	3110.74	3095.17	3108.04	3107.70

Table 2. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of C_3H_2.

Reference[2]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
777.8	867.05	799.50	782.73	755.85	767.18	768.28
885.2	981.51	906.46	901.45	887.48	889.51	889.72
888.7	998.41	909.85	899.23	881.02	892.12	891.25
974.1	1081.68	986.56	976.35	950.56	968.49	969.58
1058.1	1170.03	1078.12	1066.86	1055.50	1062.71	1063.76
1275.4	1390.84	1307.34	1285.39	1264.95	1271.48	1273.24
1594.3	1738.67	1613.96	1606.51	1583.04	1593.72	1593.46
3114.6	3281.15	3173.89	3133.04	3108.12	3119.99	3117.71

Table 3. Reported fundamental vibrational frequencies from original publications and computed values from this work of C_4.

Reference[3]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
303.3	358.89	283.91	301.51	318.55	307.00	304.83
523.0	430.20	704.14	519.84	535.77	488.41	524.10
931.9	1060.22	952.91	933.45	923.36	974.53	931.17
1003.4	1034.14	1130.52	1008.09	968.61	1000.79	1003.14
1257.1	1380.45	1320.25	1262.44	1241.16	1256.48	1255.52
1318.8	1581.62	1440.03	1322.56	1446.95	1292.80	1316.21

Table 4. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of CH_4.

Reference[4]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
1309.6	1422.04	1325.43	1317.44	1307.24	1310.39	1310.37
1530.1	1632.29	1561.80	1544.07	1532.69	1531.16	1531.91
2916.3	3052.00	2984.39	2940.31	2908.72	2916.30	2917.96
3016.7	3137.01	3089.69	3021.30	3006.56	3015.79	3012.99

Table 5. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of F_2O.

Reference[5]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
465.4	588.17	481.87	464.27	459.09	465.65	465.55
845.2	1191.08	898.54	845.61	792.83	843.60	844.94
936.8	1202.62	973.50	935.06	894.93	937.17	937.08

Table 6. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of H_2CO.

Reference[6]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
1166.1	1319.85	1201.87	1185.41	1165.07	1163.98	1163.04
1246.0	1351.54	1268.78	1263.11	1246.37	1245.55	1244.53
1499.5	1620.97	1533.13	1522.15	1495.82	1499.42	1497.91
1745.0	1974.72	1752.20	1761.60	1731.18	1743.52	1743.47
2783.5	2961.94	2854.03	2807.96	2768.12	2783.07	2783.52
2844.4	3042.92	2895.23	2855.06	2832.44	2840.91	2841.97

Table 7. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of H_2O.

Reference[7]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
1596.8	1697.70	1748.29	1616.80	1592.20	1613.19	1601.29
3662.4	3972.00	3683.77	3687.22	3601.35	3603.00	3660.15
3760.9	4060.37	3866.78	3780.26	3690.03	3748.18	3757.29

Table 8. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of HO_2^+.

Reference[7]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
1060.3	1252.48	1032.06	1058.68	1030.17	1059.34	1065.28
1379.5	1594.84	1646.61	1386.08	1345.68	1510.38	1380.77
3029.2	3362.31	2904.15	3070.67	2902.42	3050.40	3022.97

Table 9. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of HOCl.

Reference[8]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
725.0	856.77	789.42	719.77	715.40	724.17	725.37
1239.8	1360.42	1214.00	1247.26	1234.83	1241.26	1241.90
3617.2	3968.40	3711.65	3646.44	3547.54	3622.27	3615.17

Table 10. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of HOF.

Reference[5]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
898.0	1164.43	977.20	898.77	856.27	896.09	897.52
1360.1	1555.53	1375.76	1358.26	1348.71	1360.67	1361.17
3604.8	3942.57	3638.55	3616.54	3528.94	3606.86	3604.92

Table 11. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of N_2H_2.

Reference[9,10]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
1292.3	1426.70	1277.25	1303.37	1293.41	1295.50	1294.59
1317.9	1437.70	1285.29	1324.15	1317.55	1318.91	1317.71
1520.2	1693.60	1526.15	1535.17	1507.27	1514.90	1517.37
1580.1	1849.84	1605.40	1586.29	1566.79	1580.64	1579.73
3040.4	3331.94	3270.76	3065.04	2960.08	3036.06	3035.53
3117.9	3432.84	3297.81	3136.57	3065.38	3114.89	3112.46

Table 12. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of NH_2^-.

Reference[11]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
1448.2	1480.82	1427.23	1448.50	1436.68	1443.51	1461.85
3124.3	3523.66	3241.04	3148.48	3078.27	3134.21	3120.23
3193.7	3558.41	3318.60	3196.99	3138.90	3205.17	3189.07

Table 13. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of NH_3.

Reference[12]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
957.6	1015.73	986.58	1014.15	890.24	950.85	949.75
1629.2	1749.28	1644.33	1645.07	1623.93	1628.26	1628.76
3344.2	3572.56	3397.86	3358.47	3328.61	3340.80	3339.48
3452.0	3662.35	3515.83	3452.41	3423.54	3447.24	3447.07

Table 14. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of NH_4^+.

Reference[13]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
1443.7	1557.32	1449.32	1453.96	1439.98	1443.17	1444.07
1688.3	1797.19	1706.18	1704.17	1686.07	1688.91	1690.04
3237.9	3433.25	3276.70	3262.85	3208.30	3230.74	3232.91
3345.2	3536.14	3392.90	3356.11	3310.45	3338.86	3338.39

Table 15. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of PH_3. Degenerate modes are not repeated.

Reference[14]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
992.0	1082.44	1005.25	1002.90	994.48	995.43	991.50
1119.0	1212.46	1146.13	1125.40	1122.78	1124.10	1119.87
2323.5	2449.80	2408.37	2331.18	2328.35	2323.01	2318.88
2328.7	2458.80	2408.34	2333.18	2330.03	2324.10	2324.73

Table 16. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of SiF_2.

Reference[15]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
343.6	377.84	340.00	342.11	343.85	344.33	344.31
857.6	906.26	850.30	855.61	854.52	859.66	859.11
873.2	910.21	867.81	874.12	869.74	875.20	874.58

Table 17. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of SiH_4.

Reference[16]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
912.5	994.50	934.25	909.89	994.50	934.25	921.53
969.1	1028.51	996.66	972.22	1028.51	996.66	966.13
2187.9	2250.42	2244.17	2192.56	2250.42	2244.17	2191.46
2190.7	2250.45	2244.15	2193.29	2250.45	2244.15	2192.11

Table 18. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of SO_2.

Reference[15]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
515.4	604.48	474.17	512.72	516.06	515.99	515.79
1146.3	1365.47	1166.74	1138.46	1130.65	1150.31	1146.32
1349.7	1559.91	1390.56	1340.66	1331.10	1354.71	1349.74

Table 19. Reported fundamental vibrational frequencies from reference publication, and computed values from this work of SO_3.

Reference[17]	HF_TZ	MP2_TZ	CC_TZ	HF_TZ_REF	MP2_TZ_REF	CC_TZ_REF
496.5	578.04	485.19	487.08	498.03	499.39	498.50
528.3	595.56	513.93	518.86	527.47	528.15	528.13
1067.0	1226.45	1048.71	1049.91	1053.19	1076.25	1068.04
1396.4	1552.39	1401.02	1379.25	1375.86	1406.98	1397.70

References

(1) Delahaye, T.; Nikitin, A.; Rey, M.; Szalay, P. G.; Tyuterev, V. G. A new accurate ground-state potential energy surface of ethylene and predictions for rotational and vibrational energy levels. *Journal of Chemical Physics* **2014**, *141*, 104301.

(2) Dateo, C. E.; Lee, T. J. An accurate ab initio quartic force field and vibrational frequencies for cyclopropenylidene. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53*, 1065-1077.

(3) Wang, X.; Huang, X.; Bowman, J. M.; Lee, T. J. Anharmonic rovibrational calculations of singlet cyclic C4 using a new ab initio potential and a quartic force field. *The Journal of chemical physics* **2013**, *139*, 224302.

(4) Lee, T. J.; Martin, J. M. L.; Taylor, P. R. An Accurate ab initio Quartic Force Field and Vibrational Frequencies for CH4 and Isotopomers. *Journal of Chemical Physics* **1995**, *102*, 254-261.

(5) Breidung, J.; Thiel, W.; Gauss, J.; Stanton, J. F. Anharmonic force fields from analytic CCSD(T) second derivatives: HOF and F2O. *The Journal of Chemical Physics* **1999**, *110*, 3687-3696.

(6) Yachmenev, A.; Jensen, P.; Yurchenko, S.; Thiel, W. A new "spectroscopic" potential energy surface for formaldehyde in its ground electronic state. *Journal of Chemical Physics* **2011**, *134*, 244307-244307-244311.

(7) Huang, X.; Lee, T. J. A procedure for computing accurate ab initio quartic force fields: Application to HO2+ and H2O. *The Journal of chemical physics* **2008**, *129*, 044312.

(8) Peterson, K. A. Accurate ab initio near-equilibrium potential energy and dipole moment functions of HOCl and HOBr. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53*, 1051-1064.

(9) Martin, J. M. L.; Taylor, P. R. Accurate ab initio quartic force field for trans-HNNH and treatment of resonance polyads. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53*, 1039-1050.

(10) Craig, N. C.; Appiah, K. J.; Miller, C. E.; Seiden, M. V.; Varley, J. E. Reevaluation of matrix-isolation infrared spectra of the isotopologues of trans-diazene and attempts to prepare cis-diazene by photoisomerization. *Journal of Molecular Spectroscopy* **2015**, *310*, 3.

(11) Huang, X.; Lee, T. J. Accurate ab initio quartic force fields for NH[sub 2]−] and CCH[sup −] and rovibrational spectroscopic constants for their isotopologs. *The Journal of Chemical Physics* **2009**, *131*, 104301.

(12) Yurchenko, S. N.; Zheng, J.; Lin, H.; Jensen, P.; Thiel, W. Potential-energy surface for the electronic ground state of NH3 up to 20,000 cm-1 above equilibrium. *The Journal of chemical physics* **2005**, *123*, 134308.

(13) Martin, J. M. L.; Lee, T. J. Accurate ab initio quartic force field and vibrational frequencies of the NH4+ ion and its deuterated forms. *Chemical Physics Letters* **1996**, *258*, 129-135.

(14) Ovsyannikov, R. I.; Thiel, W.; Yurchenko, S. N.; Carvajal, M.; Jensen, P. Vibrational energies of PH3 calculated variationally at the complete basis set limit. *The Journal of chemical physics* **2008**, *129*, 044309.

(15) Pak, Y.; Woods, R. C. Coupled cluster calculations of the potential energy surfaces and spectroscopic constants of SiF2, PF+2, SO2, PO−2, and ClO+2. *The Journal of Chemical Physics* **1996**, *104*, 5547-5554.

(16) Martin, J. M. L.; Baldridge, K. K.; Lee, T. J. Accurate ab initio anharmonic force field and heat of formation for silane. *Molecular Physics* **1999**, *97*, 945.

(17) Martin, J. L. A fully ab initio quartic force field of spectroscopic quality for SO3. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1999**, *55*, 709-718.

Appendix C

Experimental Section and Supporting Information
Chapter 4

Contents

Experimental Section

Supporting Information

Experimental section

All the experimental work and sample preparation were carried out under dried nitrogen atmosphere using standard Schlenk techniques. CH_2Cl_2, diethylether, ethylacetate, D_2O, triethylamine and diphenylamine were obtained commercially. Solvents were dried using an in-house solvent purification system.

TGA/DSC data were collected on dried samples using a TA Instruments SDT Q600 at 10 °C min^{-1}. Mid-IR data were collected at room temperature by using a Vertex 70 FT-IR spectrometer from Bruker (Germany), operating with a Platinum ATR unit with a diamond crystal. A resolution of 4 cm^{-1} and 16 scans were taken. Microanalyses were performed by Campbell Microanalytical Laboratory, University of Otago, Dunedin.

1. Diethylaminobis(diphenylamino)cyclopropenium chloride dihydrate $[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$. A solution of $HNPh_2$ (9.87 g, 58.37 mmol) and NEt_3 (12.2 mL, 87.56 mmol) in CH_2Cl_2 (75 mL) was added slowly to a solution of C_3Cl_5H (4.14 g, 19.46 mmol) in CH_2Cl_2 (50 mL) and cooled to 0 °C. After stirring for 1h, the reaction mixture was stirred overnight at ambient temperature. CH_2Cl_2 was removed *in vacuo* to give a dark-brown liquid. ES-MS shows the presence of $[HNEt_3]^+$, $[H_2NPh_2]^+$, $[C_3(NEt_2)_2(NPh_2)]^+$ and $[C_3(NPh_2)_2(NEt_2)]^+$. $[HNEt_3]Cl$ was filtered off after precipitation by the addition of acetone (75 mL). $[H_2NPh_2]Cl$ and $[C_3(NEt_2)_2(NPh_2)]Cl$ salts were separated by column chromatography using silica-gel 60G and CH_2Cl_2-ethylacetate (1:1 v/v) as the mobile phase. Ethanol was then used to elute $[C_3(NPh_2)_2(NEt_2)]Cl$ and the solvent was removed *in vacuo* to give a pale yellow solid. Undried CH_2Cl_2-ethylacetate (1:1 v/v) was used for crystallization, and slow evaporation at ambient temperature gave pale yellow plates (3.54 g, Yield 41%).^{1}H NMR (400 MH*z*, CD_3CN): δ = 0.88 (t, ^3J(H,H) = 7.04 Hz, 2H; NCH_2CH_3), 2.76 (q, ^3J(H,H) = 7.04 Hz, 3H; NCH_2CH_3), 7.16-7.34 ppm

(m, 10H; NC_6H_5). $^{13}C(^1H)$ NMR (100 MHz, $CDCl_3$): δ = 13.54 (NCH_2CH_3), 46.77 (NCH_2CH_3), 122.31, 124.30 (C_3) ring, 114.47, 127.22, 129.66, 142.83 (NC_6H_5). EI-MS: Found m/z 445.2490 [M^+]; Calcd 445.2490 [M^+]. $C_{31}H_{34}N_3ClO_2$: C, 72.70; H, 6.65; N, 8.14; Found: C, 73.08; H, 6.46; N, 8.13.

2. **$[C_3(NPh_2)_2(NEt_2)]Cl(D_2O)_2$.** NBu_3 (1.571 g, 8.39 mmol) and $NH(CH_2CF_3)_2$ (4.002 g, 20.98 mmol) were dissolved in CH_2Cl_2 (5 mL) and added slowly to C_3Cl_4 (1.045 g, 5.87 mmol) which was dissolved in CH_2Cl_2 (15 mL) at 0 °C with stirring. The stirring was continued at ambient temperature for 48 h. CH_2Cl_2 was removed in vacuo and the residue was dissolved in deionized water (150 mL) and extracted with diethylether (4 × 50 mL). Diethylether was removed using a rotary evaporator. The product was then dissolved in $CHCl_3$ and extracted with deionized water (6 × 50 mL). Water was removed using rotary evaporator. The residue was dissolved in CH_2Cl_2 and a white precipitate was filtered off. Removal of CH_2Cl_2 gave an off-white product (1.0 g, 41%). Crystals suitable for X-ray diffraction were grown by slow evaporation of an undried CH_2Cl_2/diethylether solution. ^1H NMR (CD_3OD, 400 MHz): 4.46 (q, $^3J_{HF}$ = 8.5 Hz, 8H, NCH_2CF_3), 3.52 (t, $^3J_{HH}$ = 7.8 Hz, 4H, NCH_2), 1.68 (m, 4H, NCH_2CH_2), 1.37 (m, 4H, $NCH_2CH_2CH_2$), 0.99 (t, $^3J_{HH}$ = 7.3 Hz, 6H, CH_3). ^{13}C NMR ($CDCl_3$, 100 MHz): 124.61 (C_3-NCH_2CH_3), 123.29 (q, $^1J_{CF}$ = 278.9 Hz, NCH_2CF_3), 120.31 (C_3-NCH_2CF_3), 56.49 (q, $^2J_{CF}$ = 36.1 Hz, NCH_2CF_3), 53.04 (NCH_2), 29.68 (NCH_2CH_2), 19.61 ($NCH_2CH_2CH_2$), 13.57 (CH_3). ^{19}F NMR (CD_3OD, 376 MHz): –72.76 (t, 6F, $^3J_{HF}$ = 8.5 Hz, NCH_2CF_3), ES^+ m/z 524.1969 (100%, M^+); calculated for $C_{19}H_{26}F_{12}N_3^+$ 524.1930. Microanalysis: Found: C, 39.94; H, 4.74; N, 6.89%. Calc. for $C_{19}H_{28}N_3ClF_{12}O$: C, 39.49; H, 4.88; N, 7.27%.

3. **Crystal Data.** A suitable crystal was mounted on a SuperNova, Dual, Cu at home/near, Atlas diffractometer. Using Olex2,[2] the structure was solved with the XS structure solution

program[3] using Direct Methods and refined with the XL refinement package[3] using Least Squares minimisation with anisotropic thermal parameters for all non-hydrogen atoms. For **2**.H_2O, hydrogen atoms on the methylene groups were refined isotropically at their calculated positions and methyl groups were refined as rotating groups; the water protons were located from the density difference maps and refined isotropically. For **1**.H_2O and **3**.H_2O, all hydrogen atoms were located from the density difference maps and refined isotropically. CCDC-2003399, 2003397 and 2003398 contain the crystallographic data **1**.H_2O, **2**.H_2O and **3**.H_2O, respectively. This data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

Supporting Information

Table1S. Selected bond distances and angles for $[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$.

Bond	Distance [Å]	Angle	[°]
C3–C2	1.365(2)	C2-C3-C1	60.58(10)
C3–C1	1.388(2)	C3-C2-C1	60.53(10)
C2–C1	1.389(2)	C3-C1-C2	58.89(10)
N3–C31A	1.4434(18)	N3-C3-C2	151.00(14)
N3–C31	1.4290(18)	N3-C3-C1	148.41(14)
N3–C3	1.3469(19)	N2-C2-C3	150.15(14)
N2–C21A	1.4350(18)	N2-C2-C1	149.31(14)
N2–C21	1.4367(18)	N1-C1-C3	152.18(14)
N2–C2	1.3424(19)	N1-C1-C2	148.92(15)
N1–C1	1.3064(19)	C31-N3-C31A	121.82(12)
N1–C11	1.4696(19)	C3-N3-C31A	117.58(12)
N1–C13	1.481(2)	C3-N3-C31	120.52(12)
C26A–Cl1	3.6616(17)	C21A-N2-C21	120.25(11)
C35–Cl1	3.6410(16)	C2-N2-C21A	119.52(12)
C36A–Cl1	3.6824(17)	C2-N2-C21	120.15(12)
C12–O1	3.462(4)	C1-N1-C11	121.40(13)
C34A–O2	3.406(3)	C1-N1-C13	119.38(13)
		C11-N1-C13	119.18(12)

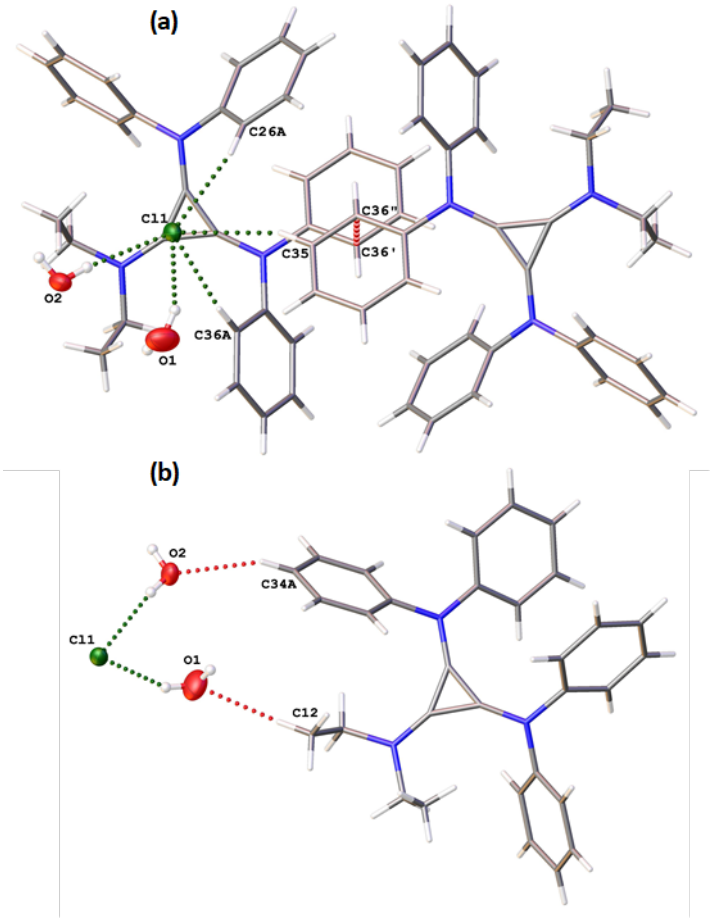


Figure 1S. Hydrogen bonding in the $[Cl_2(H_2O)_4]^{2-}$ cluster: (a) chloride environment; (b) water environment. Cl1---C26A = 3.6613(17) Å; Cl1---C35 = 3.6414(16) Å; Cl1---C36A = 3.6819(17) Å; O1---C12 = 3.457(4) Å; O2---C34A = 3.408(3) Å.

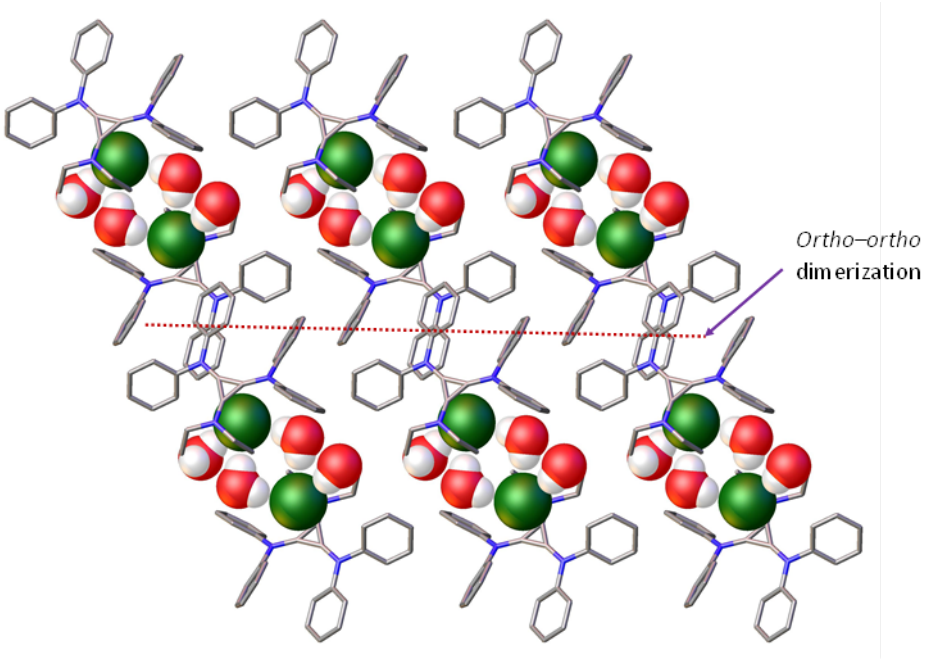


Figure 2S. Self-assembly of $[Cl_2(H_2O)_4]^{2-}$ and $[C_3(NPh_2)_2(NEt_2)]_2^{2+}$ dimer formation.

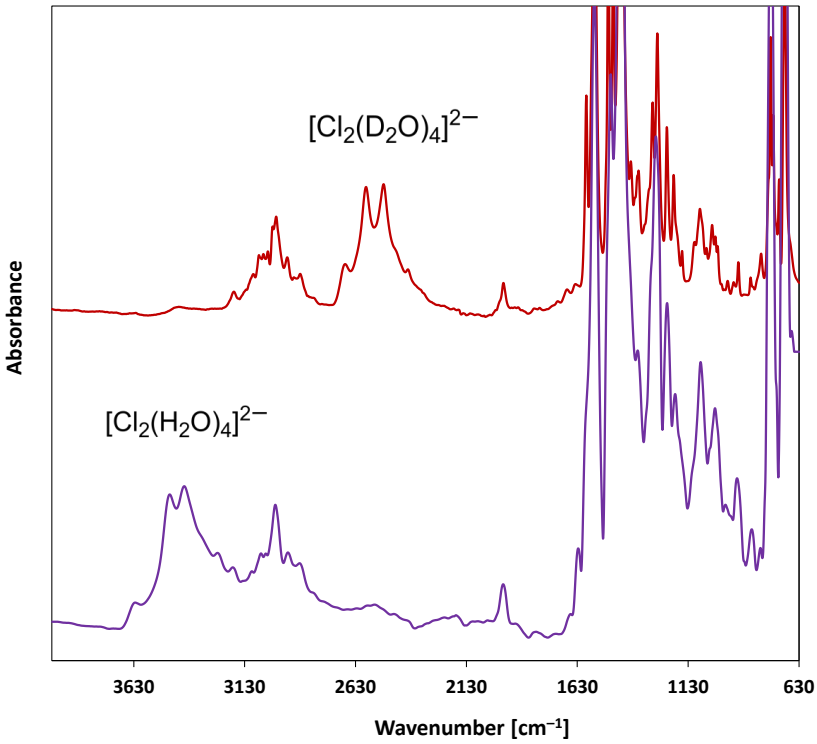


Figure 3S. FT-IR spectra of $[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$ and $[C_3(NPh_2)_2(NEt_2)]Cl.2D_2O$.

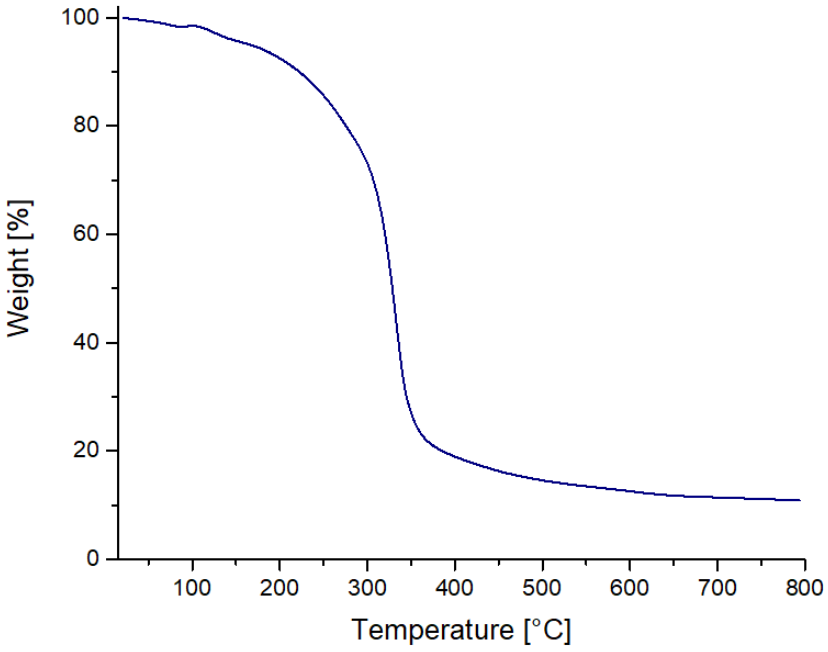


Figure 4S. TGA of the salt $[C_3(NPh_2)_2(NEt_2)]Cl.2H_2O$ at 10 °C min^{-1}

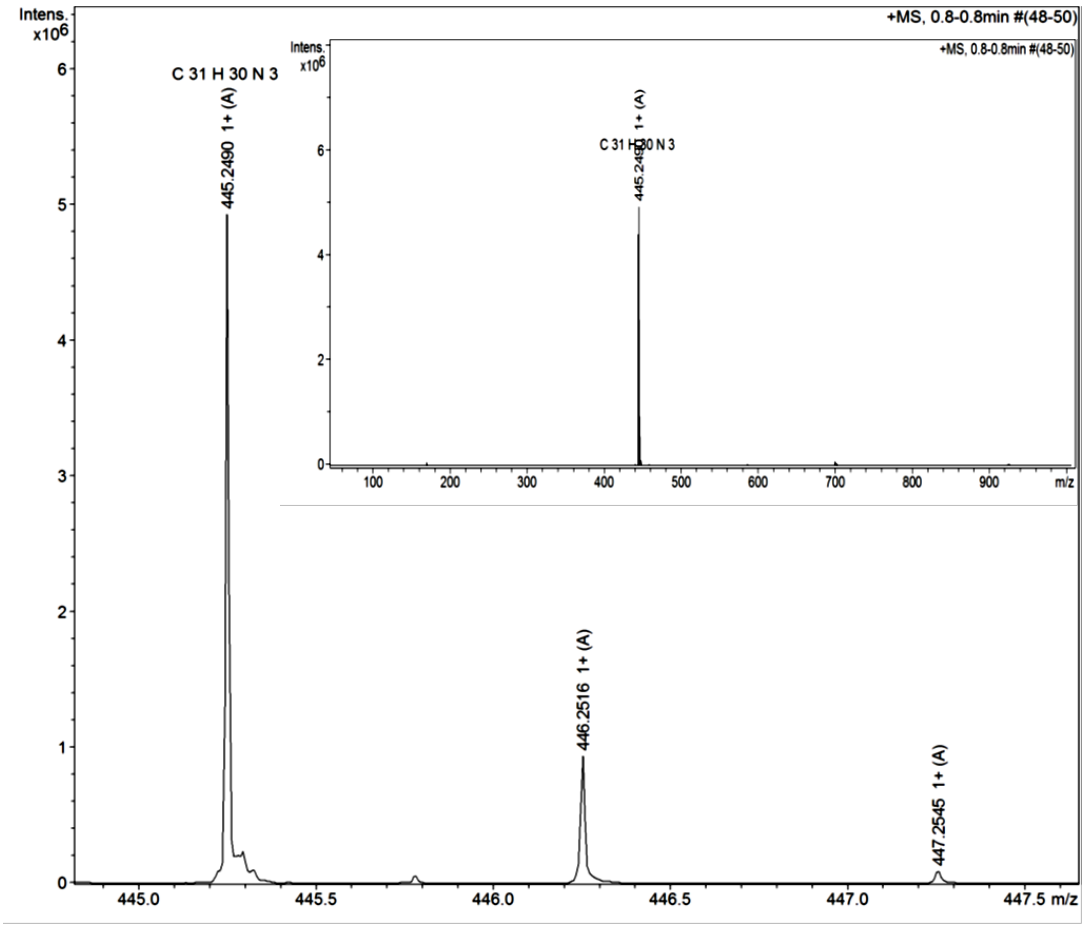


Figure 5S. ESI-MS spectrum of the $[C_3(NPh_2)_2(NEt_2)]^+$ cation

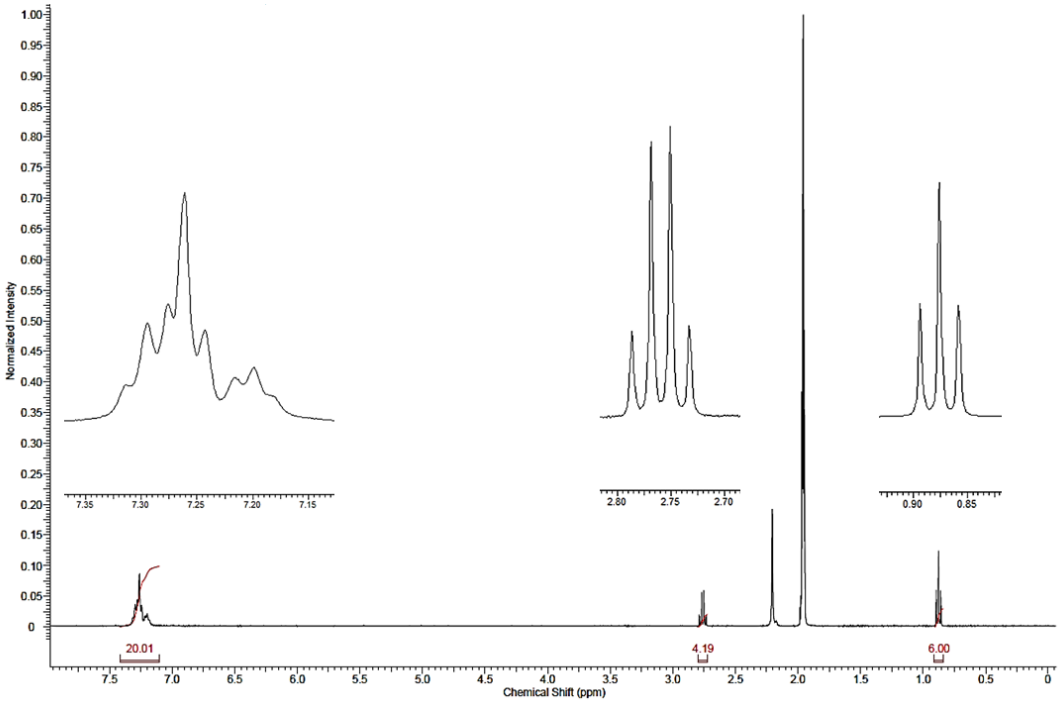


Figure 6S. ^{1}H spectrum of $[C_3(NPh_2)_2(NEt_2)]Cl(H_2O)_2$ in CD_3CN

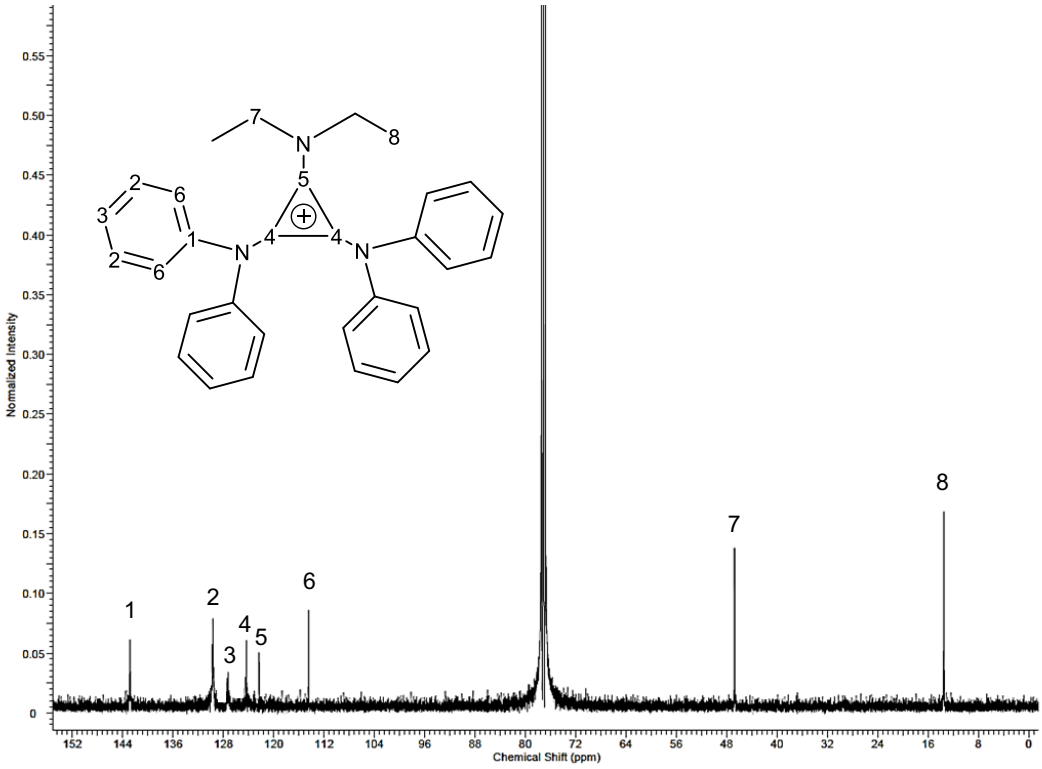


Figure 7S. ^{13}C NMR spectrum of $[C_3(NPh_2)_2(NEt_2)]Cl(H_2O)_2$ in $CDCl_3$

Appendix D

Experimental section and Supporting Information

Chapter 5

Contents

Experimental Section

Supporting Information

References

Experimental section

All the experimental work and sample preparation were carried out under dried nitrogen atmosphere using standard Schlenk techniques unless otherwise noted. CH_2Cl_2, $CHCl_3$, diethylether, ethanol, petroleum ether, D_2O, tributylamine, dibutylamine and bis(2,2,2-trifluoroethyl)amine were obtained commercially. $[C_3(NMe_2)_3]Cl$ was prepared by a literature method.[1] Solvents were dried using an in-house solvent purification system.

^1H-, $^{13}C\{^1H\}$- and ^{19}F-NMR spectra were collected on an JEOL ECZ400S spectrometer operating at 400, 100 and 376 MHz, respectively, in CD_3OD or $CDCl_3$ referenced to residual solvent peaks. The instrument was equipped with a ROYAL HFX probe. Electrospray mass spectrometry was carried out on a Micromass LCT, with samples dissolved in acetonitrile. Mid-IR data were collected at room temperature by using an Alpha-Bruker IR spectrometer, operating with a Platinum ATR unit with a diamond crystal. A resolution of 4 cm^{-1} and 16–40 scans were taken. Microanalyses were performed by Campbell Microanalytical Laboratory, University of Otago, Dunedin.

1. **Bis(2,2,2-trifluoroethyl)aminobis(dibutylamino)cyclopropenium chloride hydrate, $[C_3(NBu_2)_2(NC_2H_4F_6)]Cl.H_2O$.** NBu_3 (6.27 g, 33.5 mmol) and $NH(CH_2CF_3)_2$ (16.020 g, 84.54 mmol) were dissolved in CH_2Cl_2 (20 mL) and added slowly to C_3Cl_4 (4.190 g, 23.55 mmol) which was dissolved in CH_2Cl_2 (60 mL) at 0 °C. Stirring continued at 0 °C for 4 h and then at ambient temperature overnight. A mixture of $NHBu_2$ (3.038 g, 23.31 mmol) and NBu_3 (3.130 g, 16.71 mmol) in CH_2Cl_2 (5 mL) was added to the mixture at 0 °C and stirring continued at 0 °C for 4 h and then at RT for 48 h. CH_2Cl_2 was removed in vacuo and the residue was dissolved in deionized water (100 mL) and the product extracted with diethylether (5 × 75 mL). Diethylether was removed in vacuo. The product was dissolved in $CHCl_3$ and washed with deionized water (6 × 50 mL). $CHCl_3$ was removed in vacuo. The resultant mixture was dissolved in diethylether (50 mL) and

extracted with water (5 × 30 mL). Removal of water yielded a light-brown solid (4.2 g, 35%). Crystals suitable for X-ray diffraction were grown by slow evaporation of an undried ethanol-petroleum ether solution. ^{1}H NMR ($CDCl_3$, 400 MHz): 4.37 (q, $^{3}J_{HF}$ = 8.4 Hz, 4H, NC*H*$_2$CF$_3$), 3.32 (t, $^{3}J_{HH}$ = 8.2 Hz, 6H, NCH$_2$), 1.64 (m, 6H, NCH$_2$C*H*$_2$), 1.32 (m, 6H, NCH$_2$CH$_2$C*H*$_2$), 0.95 (t, $^{3}J_{HH}$ = 7.3 Hz, 12H, CH$_3$). ^{13}C NMR ($CDCl_3$, 100 MHz): 123.36 (q, $^{1}J_{CF}$ = 280.3 Hz, NCH$_2$*C*F$_3$), 120.33 (C$_3$-NCH$_2$CF$_3$), 116.10 (C$_3$-NBu$_2$), 57.18 (q, $^{2}J_{CF}$ = 34.7 Hz, N*C*H$_2$CF$_3$), 52.80 (NCH$_2$), 30.31 (NCH$_2$*C*H$_2$), 19.79 (NCH$_2$CH$_2$*C*H$_2$), 13.68 (CH$_3$). ^{19}F NMR (CD_3OD, 376 MHz): –72.61 (t, 6F, $^{3}J_{HF}$ = 8.2 Hz, NCH$_2$CF$_3$), ES$^+$ m/z 430.2653 (100%, M$^+$); calculated for $C_{19}H_{26}F_{12}N_3^+$ 430.2651. Melting point 102 °C. Microanalysis: Found: C, 52.25; H, 8.03; N, 7.50%. Calc. for $C_{23}H_{34}N_3F_6Cl.H_2O$: C, 52.51; H, 8.05; N, 7.99%.

2. **Bis(bis(2,2,2-trifluoroethyl)amino)dibutylaminocyclopropenium chloride hydrate, [C$_3$(NBu$_2$)(NC$_2$H$_4$F$_6$)$_2$]Cl.H$_2$O.** NBu$_3$ (1.571 g, 8.39 mmol) and NH(CH$_2$CF$_3$)$_2$ (4.002 g, 20.98 mmol) were dissolved in CH_2Cl_2 (5 mL) and added slowly to C_3Cl_4 (1.045 g, 5.87 mmol) which was dissolved in CH_2Cl_2 (15 mL) at 0 °C with stirring. The stirring was continued at ambient temperature for 48 h. CH_2Cl_2 was removed in vacuo and the residue was dissolved in deionized water (150 mL) and extracted with diethylether (4 × 50 mL). Diethylether was removed using a rotary evaporator. The product was then dissolved in $CHCl_3$ and extracted with deionized water (6 × 50 mL). Water was removed using rotary evaporator. The residue was dissolved in CH_2Cl_2 and a white precipitate was filtered off. Removal of CH_2Cl_2 gave an off-white product (1.0 g, 41%). Crystals suitable for X-ray diffraction were grown by slow evaporation of an undried CH_2Cl_2/diethylether solution. ^{1}H NMR (CD_3OD, 400 MHz): 4.46 (q, $^{3}J_{HF}$ = 8.5 Hz, 8H, NCH$_2$CF$_3$), 3.52 (t, $^{3}J_{HH}$ = 7.8 Hz, 4H, NCH$_2$), 1.68 (m, 4H, NCH$_2$C*H*$_2$), 1.37 (m, 4H, NCH$_2$CH$_2$C*H*$_2$), 0.99 (t, $^{3}J_{HH}$ = 7.3 Hz, 6H, CH$_3$). ^{13}C NMR ($CDCl_3$, 100 MHz): 124.61 (C$_3$-NCH$_2$CH$_3$), 123.29 (q, $^{1}J_{CF}$ = 278.9 Hz, NCH$_2$*C*F$_3$), 120.31 (C$_3$-NCH$_2$CF$_3$), 56.49 (q, $^{2}J_{CF}$ = 36.1 Hz, N*C*H$_2$CF$_3$), 53.04 (NCH$_2$), 29.68 (NCH$_2$*C*H$_2$), 19.61 (NCH$_2$CH$_2$*C*H$_2$), 13.57 (CH$_3$). ^{19}F NMR (CD_3OD, 376 MHz): –72.76 (t, 6F, $^{3}J_{HF}$ = 8.5 Hz, NCH$_2$CF$_3$), ES$^+$ m/z 524.1969 (100%, M$^+$);

calculated for $C_{19}H_{26}F_{12}N_3^+$ 524.1930. Microanalysis: Found: C, 39.94; H, 4.74; N, 6.89%. Calc. for $C_{19}H_{28}N_3ClF_{12}O$: C, 39.49; H, 4.88; N, 7.27%.

3. **Synthesis of deuterated and partially-deuterated samples.** A few drops of D_2O or 1:1 H_2O/D_2O were added to **2**.H_2O or **3**.H_2O (1 g) in CH_2Cl_2 (10 mL) and stirred for 10 min. The solvent was then removed *in vacuo* and the procedure repeated until the desired product was obtained, as determined by IR spectroscopy.

4. **Crystal Data.** A suitable crystal was mounted on a SuperNova, Dual, Cu at home/near, Atlas diffractometer. Using Olex2,[2] the structure was solved with the XS structure solution program[3] using Direct Methods and refined with the XL refinement package[3] using Least Squares minimisation with anisotropic thermal parameters for all non-hydrogen atoms. For **2**.H_2O, hydrogen atoms on the methylene groups were refined isotropically at their calculated positions and methyl groups were refined as rotating groups; the water protons were located from the density difference maps and refined isotropically. For **1**.H_2O and **3**.H_2O, all hydrogen atoms were located from the density difference maps and refined isotropically. CCDC-2003399, 2003397 and 2003398 contain the crystallographic data **1**.H_2O, **2**.H_2O and **3**.H_2O, respectively. This data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

Supporting Information

Table1S. Bond distances (Å) and angles (°) for **1**.H_2O.

C1–C1[1]	1.377(2)	N1–C1	1.3336(13)
C2–C1	1.3816(16)	N2–C2	1.3297(19)
N1–C11	1.4660(13)	N2–C13	1.4613(12)
N1–C12	1.4614(14)		
C1[1]-C1-C2	60.11(5)	C1-C2-C1[1]	59.77(10)
N1-C1-C1[1]	150.25(6)	N2-C2-C1	150.05(5)
N1-C1-C2	149.54(9)	C2-N2-C13	119.52(6)
C1-N1-C12	118.05(9)	C13-N2-C13	117.29(12)
C1-N1-C11	118.21(9)	C12-N1-C11	117.82(9)

Table2S. Hydrogen-bonding parameters for **1**.H_2O.

Hydrogen bond (X–H---Y)	X–H (Å)	Y---H (Å)	X---Y	X–H---Y (°)
O1–H1---Cl1	0.854(18)	2.356(18)	3.2095(9)	179.2(17)
C11–H11c---Cl1	0.987(19)	2.750(17)	3.6000(14)	144.6(12)
C12–H12a---Cl1	0.973(15)	3.101(17)	3.9551(11)	147.2(10)
C13–H13a---Cl1	0.993(18)	2.915(18)	3.8778(13)	163.8(13)
C13–H13c---Cl1	0.966(19)	2.902(18)	3.6637(14)	136.5(12)
C11–H11b---O1	0.973(16)	2.613(17)	3.5137(17)	154.0(15)
C12–H12b---O1	0.975(19)	2.603(19)	3.5059(15)	154.1(14)

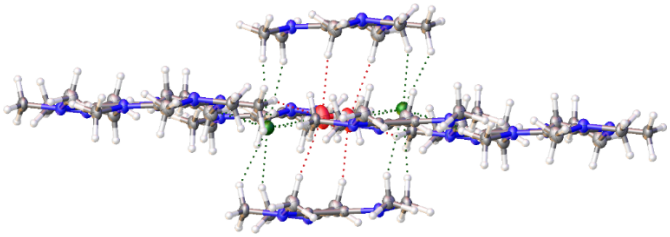


Figure 1S. Side view of the hydrogen bonding in **1**.H_2O.

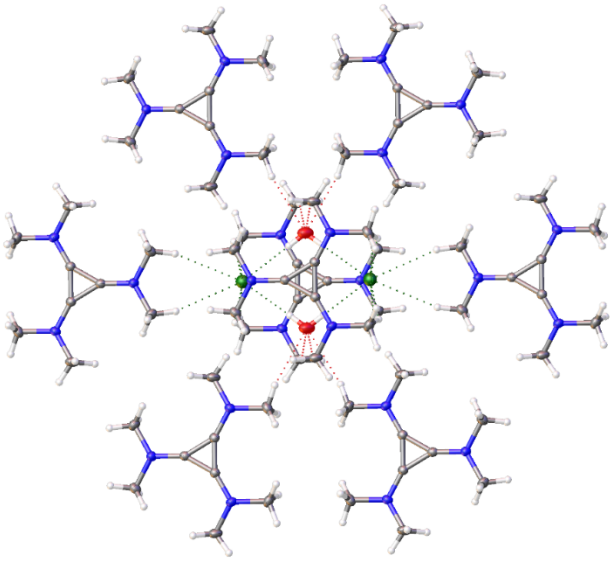


Figure 2S. Top view of the hydrogen bonding in **1**.H_2O.

Table 3S. Bond distances (Å) for **2**.H_2O.

C2 C1	1.380(2)	N1 C1	1.350(2)
C3 C1	1.379(2)	N2 C2	1.320(2)
C3 C2	1.394(2)	N3 C3	1.328(2)
N1 C11	1.456(2)	N1 C13	1.448(2)
N2 C21	1.476(2)	N2 C25	1.475(2)
N3 C31	1.471(2)	N3 C35	1.466(2)
C11 C12	1.505(2)	C13 C14	1.508(2)
F11 C12	1.332(2)	F14 C14	1.338(2)
F12 C12	1.342(2)	F15 C14	1.335(2)
F13 C12	1.344(2)	F16 C14	1.333(2)
C21 C22	1.517(2)	C26 C25	1.519(2)
C22 C23	1.522(2)	C26 C27	1.525(2)
C23 C24	1.528(3)	C27 C28	1.521(3)
C31 C32	1.525(3)	C35 C36	1.517(3)
C32 C33	1.527(2)	C36 C37	1.526(3)
C33 C34	1.525(3)	C37 C38	1.517(4)

Table 4S. Bond angles (°) for **2**.H_2O.

C3 C1 C2	60.69(11)	N1 C1 C2	150.21(15)
C1 C2 C3	59.59(11)	N1 C1 C3	149.11(16)
C1 C3 C2	59.72(11)	N2 C2 C1	151.73(15)
N3 C3 C1	151.69(16)	N2 C2 C3	148.64(15)
N3 C3 C2	148.59(16)	C2 N2 C21	121.66(14)
C1 N1 C11	118.09(13)	C2 N2 C25	119.03(13)
C1 N1 C13	119.52(14)	C25 N2 C21	118.01(13)
C13 N1 C11	120.21(13)	C3 N3 C35	122.27(14)
C3 N3 C31	119.72(14)	C35 N3 C31	117.99(13)
N1 C11 C12	112.48(15)	N1 C13 C14	112.61(14)
F11 C12 C11	114.10(14)	F14 C14 C13	110.49(15)
F12 C12 C11	110.43(16)	F15 C14 C13	113.53(16)
F13 C12 C11	112.26(14)	F16 C14 C13	112.56(14)
F11 C12 F12	106.60(15)	F15 C14 F14	106.06(15)
F11 C12 F13	106.13(16)	F16 C14 F14	107.19(17)
F12 C12 F13	106.89(14)	F16 C14 F15	106.58(16)
N2 C21 C22	112.75(14)	N2 C25 C26	113.01(14)
C21 C22 C23	113.93(14)	C25 C26 C27	111.16(14)
C22 C23 C24	112.25(16)	C28 C27 C26	112.60(16)
N3 C31 C32	112.57(14)	N3 C35 C36	113.40(15)
C31 C32 C33	111.48(15)	C35 C36 C37	111.08(17)
C34 C33 C32	112.33(16)	C38 C37 C36	113.2(2)

Table 5S. Hydrogen-bonding parameters for **2**.H_2O.

Hydrogen bond (X–H---Y)	X–H (Å)	Y---H (Å)	X---Y	X–H---Y (°)
O1–H1B---Cl1	0.83(3)	2.39(3)	3.2568(18)	170(3)
O1–H1A---Cl1[1]	0.85(4)	2.41(4)	3.2182(15)	176(3)
C11-H11B-Cl1	0.9700(17)	2.7821(4)	3.6093(16)	143.62(11)
C13-H13A-Cl1	0.9701(16)	2.6636(4)	3.5840(17)	158.53(11)
C13-H13B-Cl1[1]	0.9700(17)	2.8191(5)	3.5793(19)	135.84(11)
C21-H21A-Cl1[1]	0.9700(16)	2.8243(4)	3.5727(18)	134.59(12)
C25-H25B-Cl1	0.9700(16)	3.0357(4)	3.7309(16)	129.71(9)
C21-H21B-Cl1	0.9700(18)	2.8298(5)	3.7532(19)	159.39(9)
C11-H11A-O1	0.970(2)	2.5866(16)	3.120(3)	114.78(10)

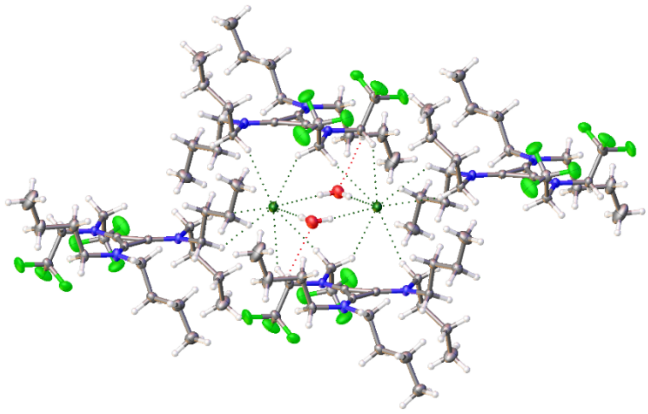


Figure 3S. Hydrogen bonding in **2**.H_2O.

Table 6S. Bond distances (Å) for **3**.H_2O.

C2	C1	1.3723(19)	N1	C1	1.3421(18)
C3	C1	1.390(2)	N2	C2	1.3434(18)
C2	C3	1.3859(19)	N3	C3	1.3196(18)
N1	C11	1.4546(19)	N1	C13	1.4566(19)
C11	C12	1.505(2)	C13	C14	1.500(2)
F11	C12	1.330(2)	F14	C14	1.330(2)
F12	C12	1.327(2)	F15	C14	1.317(2)
F13	C12	1.324(2)	F16	C14	1.336(2)
N2	C21	1.4534(18)	N2	C23	1.4525(17)
C21	C22	1.501(2)	C23	C24	1.509(2)
F21	C22	1.334(2)	F24	C24	1.3264(18)
F22	C22	1.331(2)	F25	C24	1.3371(18)
F23	C22	1.334(2)	F26	C24	1.3454(19)
N3	C31	1.4800(18)	N3	C35	1.4733(17)
C31	C32	1.5223(19)	C35	C36	1.521(2)
C32	C33	1.5241(19)	C36	C37	1.524(2)
C33	C34	1.523(2)	C37	C38	1.514(2)

Table 7S. Bond angles (°) for **3**.H_2O.

C2 C1 C3	60.23(10)	N1 C1 C2	150.99(14)
C1 C2 C3	60.51(10)	N1 C1 C3	148.77(13)
C2 C3 C1	59.26(10)	N3 C3 C1	152.00(13)
N2 C2 C1	150.72(14)	N3 C3 C2	148.72(13)
N2 C2 C3	148.72(13)	C2 N2 C21	121.39(12)
C1 N1 C11	118.44(12)	C2 N2 C23	119.95(12)
C1 N1 C13	120.14(12)	C23 N2 C21	118.63(11)
C11 N1 C13	118.77(12)	C3 N3 C35	121.84(12)
C3 N3 C31	119.34(11)	C35 N3 C31	118.67(11)
N1 C11 C12	111.74(14)	N1 C13 C14	112.90(13)
F11 C12 C11	112.44(13)	F14 C14 C13	112.56(14)
F12 C12 C11	110.50(17)	F15 C14 C13	112.83(15)
F13 C12 C11	113.23(15)	F16 C14 C13	110.21(16)
F12 C12 F11	106.12(15)	F15 C14 F14	107.76(19)
F13 C12 F11	105.66(17)	F14 C14 F16	105.67(14)
F13 C12 F12	108.51(17)	F15 C14 F16	107.43(16)
N2 C21 C22	111.09(13)	N2 C23 C24	112.41(12)
F21 C22 C21	112.62(13)	F24 C24 C23	113.58(12)
F22 C22 C21	111.19(15)	F25 C24 C23	109.29(13)
F23 C22 C21	112.11(13)	F26 C24 C23	112.48(12)
F22 C22 F21	108.37(15)	F24 C24 F25	107.94(12)
F21 C22 F23	105.72(16)	F24 C24 F26	106.28(13)
F22 C22 F23	106.46(14)	F25 C24 F26	106.96(13)
N3 C31 C32	110.90(11)	N3 C35 C36	114.28(12)
C31 C32 C33	113.62(12)	C35 C36 C37	111.22(13)
C34 C33 C32	111.37(13)	C38 C37 C36	112.52(14)

Table 8S. Hydrogen-bonding parameters for **3**.H_2O.

Hydrogen bond (X–H---Y)	X–H (Å)	Y---H (Å)	X---Y	X–H---Y (°)
O1-H1-Cl1	0.87(2)	2.34(2)	3.2051(15)	175(2)
O2-H2-Cl1	0.85(2)	2.46(2)	3.2975(15)	168(2)
C13-H13A-Cl1	0.97(2)	2.72(2)	3.6208(17)	153.1(17)
C13-H13B-Cl1[1]	0.96(2)	2.90(2)	3.6929(16)	141.0(17)
C21-H21B-Cl1	0.96(2)	2.90(2)	3.5554(17)	126.7(14)
C23-H23B-Cl1	0.93(2)	2.89(2)	3.6847(14)	144.5(15)
C21-H21A-Cl1	0.95(2)	2.85(2)	3.7196(14)	153.6(16)
C35-H35B-O2	1.00(2)	2.82(2)	3.630(2)	138.8(16)
C31-H31A-O2	0.97(2)	2.84(2)	3.734(2)	152.8(15)

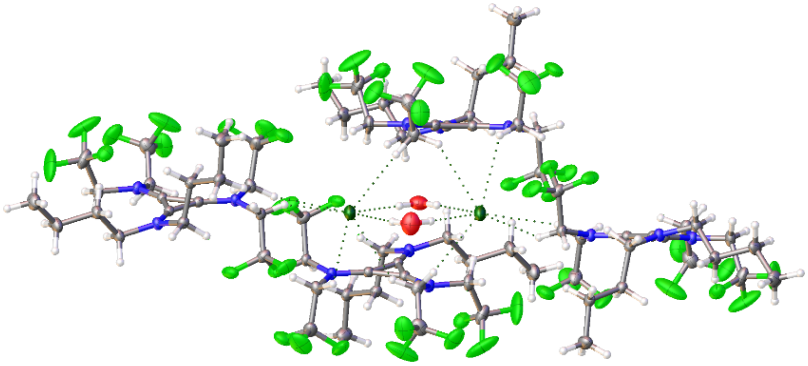


Figure 4S. View down C_2 axis (O---O) of the dichloride hydrate in **3**.H_2O. The cations in front and behind are not shown.

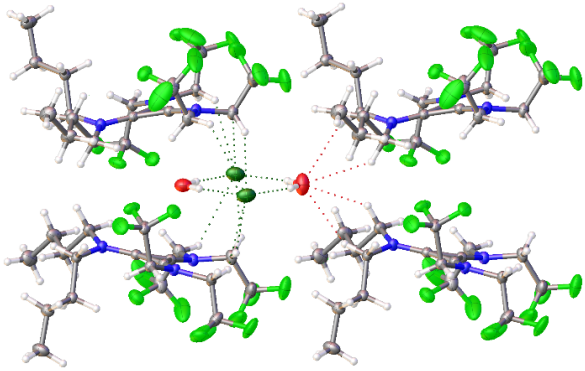


Figure 5S. View down the Cl---Cl axis of the dichloride hydrate in **3**.H_2O. The cations in front and behind are not shown.

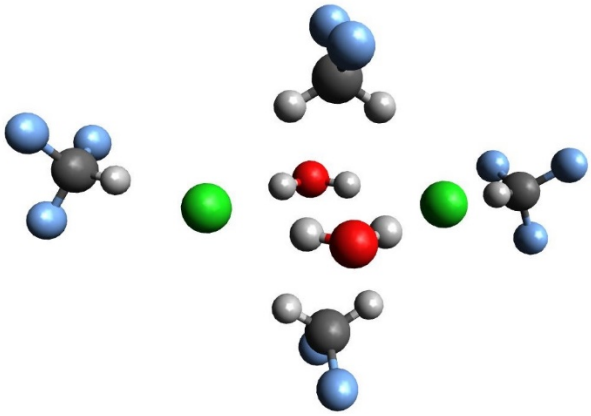


Figure 6S. Calculated structure of $[Cl_2(H_2O)_2]^{2-}.2CHF_3.2CH_2F_2$.

Table 9S. Calculated geometry of $[Cl_2(H_2O)_2]^{2-}.2CHF_3.2CH_2F_2$ at B3LYP/6-31+G*.

Cl	−2.5789751913	0.1790234466	−0.7269647040
H	−0.7706960182	0.0522325738	−2.4308308216
O	0.0047805256	−0.0079700205	−3.0235609004
H	0.7808465540	−0.0725037872	−2.4318034064
Cl	2.5878761436	−0.2069280896	−0.7419719749
H	−0.7608460602	0.0760000973	0.9380828192
O	0.0142944242	0.0195123192	1.5321772528
H	0.7837876153	−0.0589378730	0.9332148031
H	0.7220860207	−2.2747836281	−1.0652501342
C	−0.2419853574	−2.7837818620	−1.0919559572
H	−1.1106646076	−2.1293024405	−1.0135735588
F	−0.3342110650	−3.5184521628	−2.2644472602
F	−0.2840350630	−3.7061909135	−0.0541965228
H	−0.7220124999	2.2589489867	−1.0592612872
C	0.2384300188	2.7729110923	−1.1102235971
H	1.1119814251	2.1231701902	−1.0499464072
F	0.2989011301	3.5037160991	−2.2872651501
F	0.3004374879	3.6992386989	−0.0773101332
H	4.4955544461	−0.1268865186	0.6564760739
C	5.4111348140	−0.0845362008	1.2619872772
F	5.1811406366	−0.3918911370	2.5637649482
F	5.9755008300	1.1523595873	1.2393535240
F	6.3549523306	−0.9560554331	0.8108036514
H	−4.5004221358	0.0901895674	0.6516364371
C	−5.4156327292	0.0892739917	1.2592531688
F	−5.1754805228	−0.2564805839	2.5492951642
F	−6.0074579508	1.3138361631	1.2794950895
F	−6.3414379252	−0.7866679577	0.7804936727

Table 10S. Calculated vibrational frequencies of $[Cl_2(H_2O)_2]^{2-}$ and its isotopomers at B3LYP/6-31+G*.

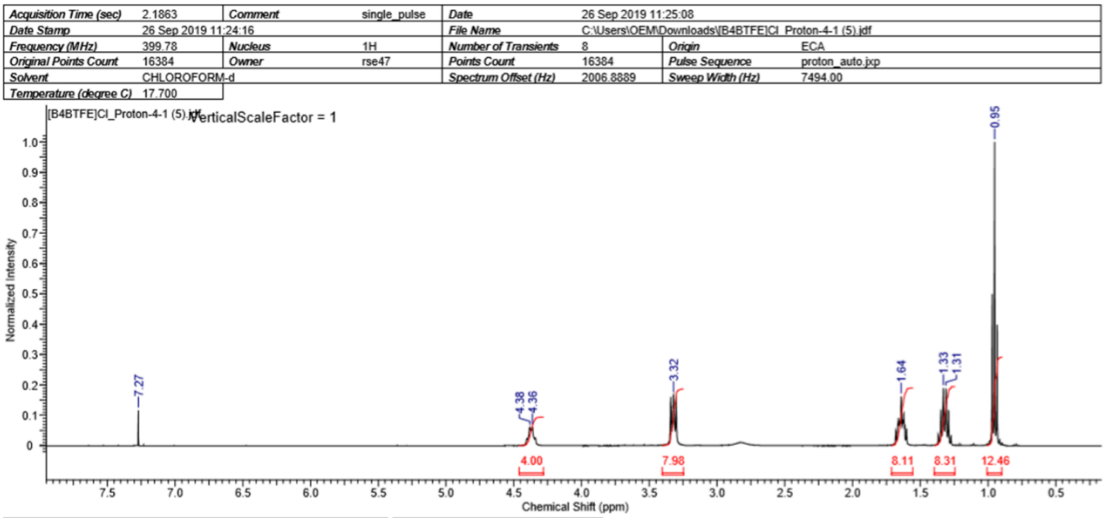
$(H_2O)_2$	Harm.	Anharm.	$(D_2O)_2$	Harm.	Anharm.
bend, anti	1698.3	1663.2	bend, anti	1240.1	1221.4
bend,syn	1717.1	1681.3	bend,syn	1253.7	1234.6
sym, anti	3577.6	3422.1	sym, anti	2580.2	2499.3
sym, syn	3582.2	3426.3	sym, syn	2583.3	2502.2
asym, anti	3662.1	3499.1	asym, anti	2686.2	2598.5
asym, syn	3679.9	3515.4	asym, syn	2699.2	2610.7

$(H_2O)(HDO)$	Harm.	Anharm.	$(HDO)(H_2O)$	Harm.	Anharm.
bend, HDO	1500.0	1472.7	bend, HDO	1502.4	1475.0
bend, H_2O	1709.5	1674.0	bend, H_2O	1706.7	1671.3
O-D str	2637.4	2552.9	O-D str	2634.3	2550.0
sym str H_2O	3580.2	3423.5	sym str H_2O	3579.4	3423.8
O-H str	3626.8	3467.0	O-H str	3624.8	3465.1
asym str H_2O	3668.3	3504.8	asym str H_2O	3675.3	3511.2

$(D_2O)(HDO)$	Harm.	Anharm.	$(HDO)(D_2O)$	Harm.	Anharm.
bend, D_2O	1247.4	1228.5	bend, D_2O	1246.0	1227.1
bend, HDO	1500.6	1473.2	bend, HDO	1503.0	1475.5
sym str D_2O	2582.2	2501.2	sym str D_2O	2581.3	2500.3
O-D str	2637.2	2552.7	O-D str	2634.1	2549.8
asym str D_2O	2690.0	2602.0	asym str D_2O	2695.8	2607.5
O-H str	3627.7	3467.8	O-H str	3625.5	3465.8

(HDO)(HDO)	Harm.	Anharm.	(HDO)(DHO)	Harm.	Anharm.
bend, anti	1493.3	1466.2	bend, syn	1493.4	1466.3
bend,syn	1509.9	1482.2	bend, anti	1509.7	1482.0
O-D str, anti	2632.2	2548.0	O-D, syn	2633.6	2549.3
O-D str, syn	2639.6	2555.0	O-D, anti	2638.1	2553.6
O-H str, anti	3621.0	3461.6	O-H, syn	3623.0	3463.5
O-H str, syn	3632.1	3471.9	O-H, anti	3630.2	3470.1

$(H_2O)(D_2O)$	Harm.	Anharm.	$(D_2O)(H_2O)$	Harm.	Anharm.
bend, D_2O	1246.0	1227.2	bend, D_2O	1247.5	1228.6
bend,H_2O	1709.3	1673.8	bend,H_2O	1706.4	1671.1
sym str D_2O	2581.3	2500.4	sym str D_2O	2582.2	2501.2
asym str D_2O	2695.6	2607.3	asym str D_2O	2689.7	2601.8
sym str H_2O	3580.3	3424.5	sym str H_2O	3579.5	3423.8
asym str H_2O	3667.5	3504.0	asym str H_2O	3674.6	3510.5



Acquisition Time (sec)	2.1863	Comment	single_pulse	Date	26 Sep 2019 11:25:08		
Date Stamp	26 Sep 2019 11:24:16			File Name	C:\Users\OEM\Downloads\[B4BTFE]Cl_Proton-4-1 (5).jdf		
Frequency (MHz)	399.78	Nucleus	1H	Number of Transients	8	Origin	ECA
Original Points Count	16384	Owner	rse47	Points Count	16384	Pulse Sequence	proton_auto.jxp
Solvent	CHLOROFORM-d			Spectrum Offset (Hz)	2006.8889	Sweep Width (Hz)	7494.00
Temperature (degree C)	17.700						

No.	(ppm)	Value	Absolute Value	Non-Negative Value
1	[0.8973 .. 1.00	2.4608354	3.35746813e+5	12.46083546
2	[1.2449 .. 1.39	8.30773258	2.23844906e+5	8.30773258
3	[1.5560 .. 1.71	8.10584736	2.18405281e+5	8.10584736
4	[3.2484 .. 3.40	7.97559214	2.14895672e+5	7.97559214
5	[4.2821 .. 4.46	4.00050545	1.07790281e+5	4.00050545

No.	(ppm)	(Hz)	Height
1	0.95	380.5	1.0000
2	1.31	523.7	0.1880
3	1.33	531.5	0.1895
4	1.64	655.9	0.1626
5	3.32	1327.8	0.1673
6	4.36	1743.6	0.0612
7	4.38	1752.3	0.0596
8	7.27	2906.4	0.1162

Figure 7S. ^{1}H-NMR spectrum of **2**.H_2O.

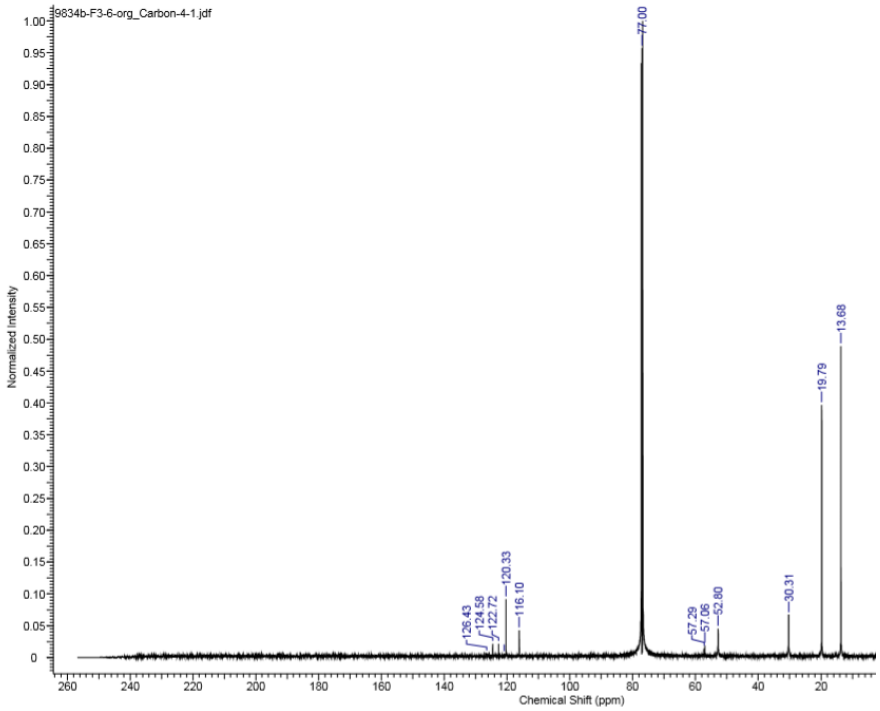
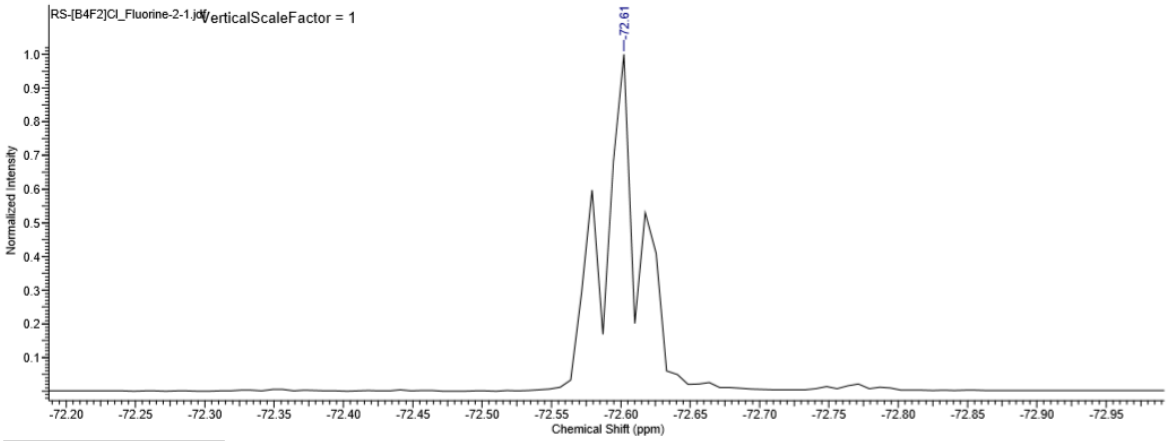


Figure 8S. $^{13}C\{^{1}H\}$-NMR spectrum of **2**.H_2O.

Acquisition Time (sec)	0.3460	*Comment*	single_pulse	*Date*	09 Mar 2020 09:43:19		
Date Stamp	09 Mar 2020 09:42:20			*File Name*	C:\Users\OEM\Downloads\RS-[B4F2]Cl_Fluorine-2-1.jdf		
Frequency (MHz)	376.17	*Nucleus*	19F	*Number of Transients*	16	*Origin*	ECA
Original Points Count	32768	*Owner*	rse47	*Points Count*	32768	*Pulse Sequence*	single_pulse.jxp
Solvent	METHANOL-d4			*Spectrum Offset (Hz)*	-37617.1055	*Spectrum Type*	STANDARD
Sweep Width (Hz)	94696.98	*Temperature (degree C)*	18.800				



No.	(ppm)	(Hz)	Height
1	-72.61	-27312.8	1.0000

Figure 9S. ^{19}F-NMR spectrum of **2**.H_2O.

School of Physical and Chemical Sciences

UC UNIVERSITY OF CANTERBURY

Mass Spectrum SmartFormula Report

Meas. m/z	#	Formula	Score	m/z	|err| [ppm]	err [mDa]	mSigma	rdb	e⁻ Conf	N-Rule
472.3153	1	C 22 H 40 F 4 N 3 O 3	0.00	470.3000	19.7	-9.3	148.0	2.5	even	ok
	2	C 23 H 39 F 6 N 3	0.00	471.3043	16.6	-7.9	153.8	3.0	odd	ok
	3	C 23 H 40 F 6 N 3	100.00	472.3121	6.7	-3.2	218.4	2.5	even	ok

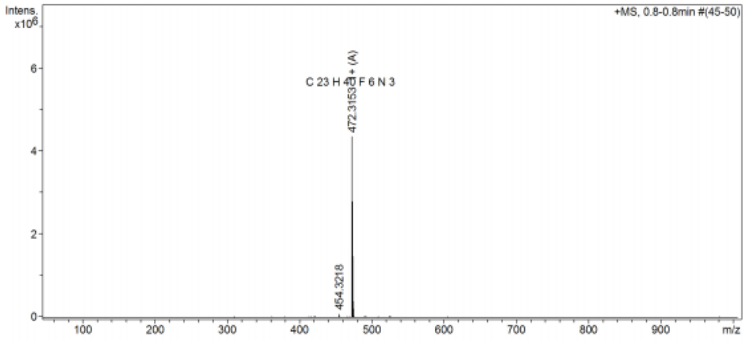
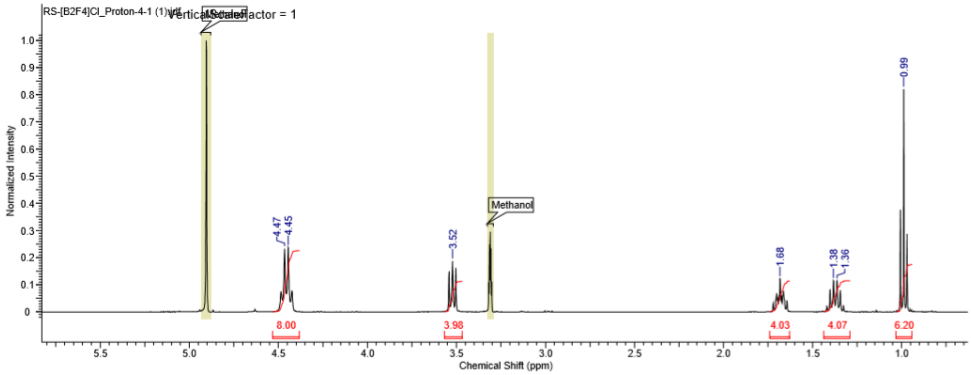


Figure 10S. EI-MS$^+$ spectrum of **2**.H_2O.

Acquisition Time (sec)	2.1863	*Comment*	single_pulse	*Date*	09 Mar 2020 09:48:01		
Date Stamp	09 Mar 2020 09:47:10			*File Name*	C:\Users\OEM\Downloads\RS-[B2F4]Cl_Proton-4-1 (1).jdf	*Frequency (MHz)*	399.78
Nucleus	1H	*Number of Transients*	8	*Origin*	ECA	*Original Points Count*	16384
Owner	rse47	*Points Count*	16384	*Pulse Sequence*	proton_auto.jxp	*Solvent*	METHANOL-d4
Spectrum Offset (Hz)	2009.6462	*Sweep Width (Hz)*	7494.00	*Temperature (degree C)*	18.600		



No.	(ppm)	Annotation	Layer No.	Created By	Created At	Modified By	Modified At
1	[3.29 .. 3.33]	Methanol	1	OEM	Mon 9/03/2020 7:14:02 PM		
2	[4.88 .. 4.93]	Methanol	1	OEM	Mon 9/03/2020 7:14:02 PM		

No.	(ppm)	Value	Absolute Value	Non-Negative Value
1	[0.9433 .. 1.03	6.19606924	1.49605078e+5	6.19606924
2	[1.2904 .. 1.434	4.07267185	9.83353125e+4	4.07267189
3	[1.5289 .. 1.744	4.03291321	9.73753281e+4	4.03291321
4	[3.4679 .. 3.563	3.97560000	9.59914922e+4	3.97560000
5	[4.8831 .. 4.53	8.00044918	1.93172125e+5	8.00044918

No.	(ppm)	(Hz)	Height
1	0.99	395.2	0.8205
2	1.36	544.7	0.1140
3	1.38	553.0	0.1155
4	1.68	672.8	0.1227
5	3.52	1408.4	0.1871
6	4.45	1777.0	0.2370
7	4.47	1785.3	0.2333

Figure 11S. ^{1}H-NMR spectrum of **3**.H_2O.

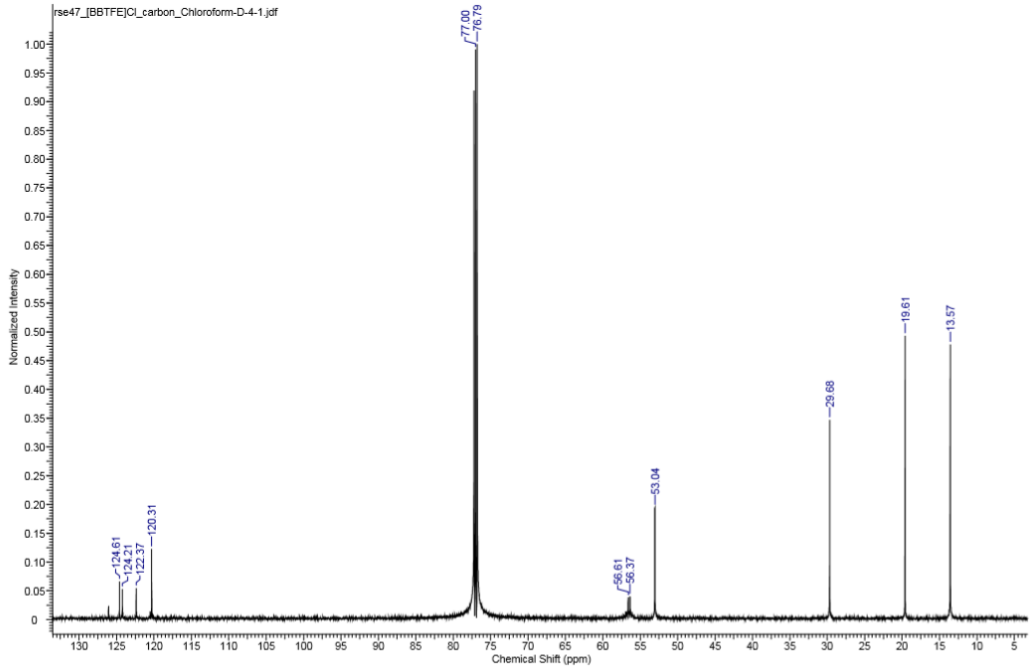


Figure 12S. $^{13}C\{^{1}H\}$-NMR spectrum of **3**.H_2O.

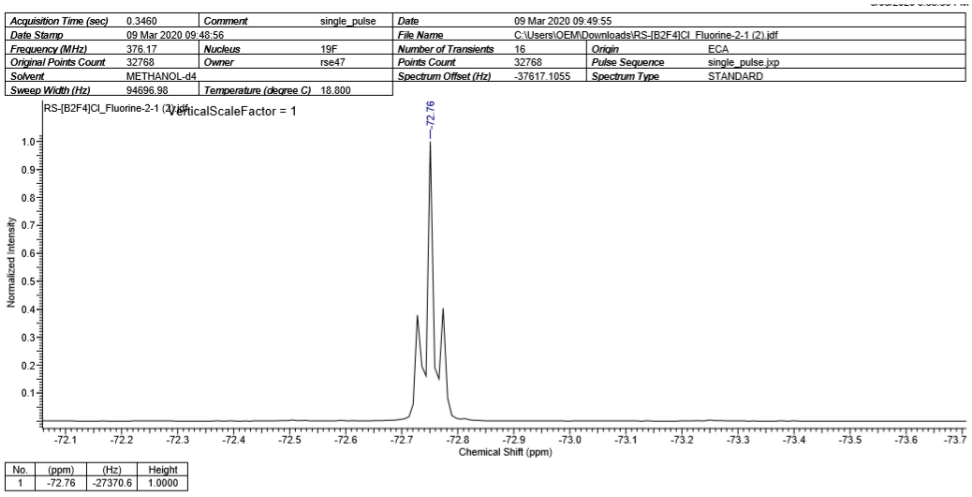


Figure 13S. ^{19}F-NMR spectrum of **3**.H_2O.

School of Physical and Chemical Sciences

UNIVERSITY OF CANTERBURY

Mass Spectrum SmartFormula Report

Meas. m/z	#	Formula	Score	m/z	|err| [ppm]	err [mDa]	mSigma	rdb	e¯ Conf	N-Rule
524.1969	1	C 19 H 26 F 12 N 3	100.00	524.1930	7.4	-3.9	216.4	2.5	even	ok

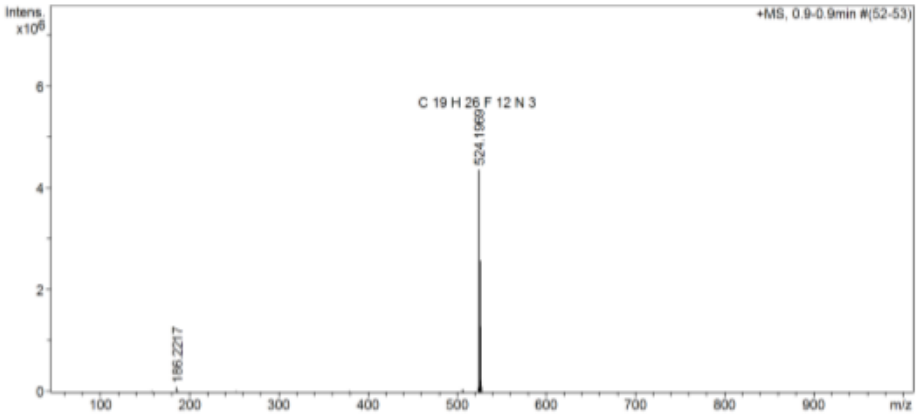


Figure 14S. EI-MS$^+$ spectrum of **3**.H_2O.

References

(1) Walst, K. J.; Yunis, R.; Bayley, P. M.; MacFarlane, D. R.; Ward, C. J.; Wang, R.; Curnow, O. J. Synthesis and physical properties of tris (dialkylamino) cyclopropenium bistriflamide ionic liquids. *RSC Advances* **2015**, *5*, 39565-39579.

(2) Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A.; Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography* **2009**, *42*, 339-341.

(3) Hübschle, C. B.; Sheldrick, G. M.; Dittrich, B. ShelXle: a Qt graphical user interface for SHELXL. *Journal of Applied Crystallography* **2011**, *44*, 1281-1284.

(4) Sibaev, M.; Crittenden, D. L. Quadratic corrections to harmonic vibrational frequencies outperform linear models. *The Journal of Physical Chemistry A* **2015**, *119*, 13107-13112.

Appendix E

MD Control, Molecular Structure, Topology and Force Field Files

Contents

Supporting Information

Supporting Information

1. Minimization mdp file – step 1.

```
-------------------------
Filename: minimize_1.mdp
-------------------------
; Minimization algorithm control parameters
integrator = steep
emtol = 10.0
emstep = 0.01
nsteps = 50000
; Potential energy evaluation parameters
nstlist = 1
cutoff-scheme = Verlet
ns-type = grid
periodic-molecules = yes
coulombtype = PME-Switch
coulomb-modifier = Potential-shift
rcoulomb-switch = 0.95
rcoulomb = 1.0
vdwtype = Cut-off
vdw-modifier = Force-switch
rvdw-switch = 0.95
rvdw = 1.0
rlist = 1.2
pbc = xyz
```

2. Minimization mdp files – steps 2,3,4,5

```
--------------------------------
Filenames: minimize_2/3/4/5.mdp
--------------------------------
; Minimization algorithm control parameters
integrator = cg
emtol = 1.0    ; 0.1    ; 0.01   ; 0.001
emstep = 0.001 ; 0.0001 ; 0.0001 ; 0.00001
nsteps = 50000
; Potential energy evaluation parameters
nstlist = 1
cutoff-scheme = Verlet
ns-type = grid
periodic-molecules = yes
coulombtype = PME-Switch
coulomb-modifier = Potential-shift
rcoulomb-switch = 0.95
rcoulomb = 1.0
vdwtype = Switch
vdw-modifier = Potential-shift
rvdw-switch = 0.95
rvdw = 1.0
rlist = 1.2
pbc = xyz
```

3. Minimization mdp file – step 6.

```
-------------------------
Filename: minimize_6.mdp
-------------------------
; Minimization algorithm control parameters
integrator = l-bfgs
emtol = 0.0001
emstep = 0.00001
nsteps = 50000
; Potential energy evaluation parameters
nstlist = 1
cutoff-scheme = Verlet
ns-type = grid
periodic-molecules = yes
coulombtype = PME-Switch
coulomb-modifier = Potential-shift
rcoulomb-switch = 0.95
rcoulomb = 1.0
vdwtype = Switch
vdw-modifier = Potential-shift
rvdw-switch = 0.95
rvdw = 1.0
rlist = 1.2
pbc = xyz
```

4. Equilibration mdp file – NPT ensemble.

```
-----------------------------------------
Filename: NPT.mdp (equilibration settings)
-----------------------------------------
title            =  heating phase
cpp              =  /usr/bin/cpp
constraints      =  none
integrator       =  md
dt               =  0.001    ; ps = 1 fs
nsteps           =  500000 ; initial equilibration run
;nsteps          =  50000  ; subsequent re-equilibration runs
nstcomm          =  1
nstxtcout        = 1000
nstxout          =  100000
nstvout          =  100000
nstfout          =  0
nstlog           =  1000
nstenergy        =  1000
nstlist          =  10
ns_type          =  grid
pbc              =  xyz
periodic_molecules  =  yes
coulombtype      =  PME
rlist            =  1.2
rcoulomb         =  1.0
rvdw             =  1.0
energygrps       =  system
DispCorr         =  EnerPres
optimize_fft     = yes
cutoff-scheme    = verlet
; Berendsen temperature coupling is on in two groups
Tcoupl           =  v-rescale
tau_t            =  0.1
tc-grps          =  system
ref_t            =  100
annealing        =  single
annealing_npoints   =  2
annealing_time      =  0 150
annealing_temp      =  0 100
; Pressure coupling is on
compressibility     =  2.5e-5 2.2e-5 7.9e-7 1.2e-5 4.4e-5  4.2e-5
pcoupltype       =  anisotropic
ref_p            =  1.0 1.0 1.0 1.0 1.0 1.0
Pcoupl           = berendsen
tau_p            =  1
unconstrained_start =  yes
```

```
; Generate velocites is off
gen_vel           = no
gen_seed          =  173529
```

5. Simulation mdp file – NVE ensemble.

```
-----------------------------------------
Filename: NVE.mdp (sampling settings)
-----------------------------------------
title            =  heating phase
cpp              =  /usr/bin/cpp
constraints      =  none
integrator       =  md
dt               =  0.001    ; ps = 1 fs
nsteps           =  30000
nstcomm          =  1
nstxtcout        =  1
nstxout          =  1
nstvout          =  1
nstfout          =  1
nstlog           =  1000
nstenergy        =  1000
nstlist          =  10
ns_type          =  grid
pbc              =  xyz
periodic_molecules  =  yes
coulombtype      =  PME
rlist            =  1.2
rcoulomb         =  1.0
rvdw             =  1.0
energygrps       =  system
DispCorr         =  EnerPres
optimize_fft     = yes
cutoff-scheme    = verlet
; Temperature coupling is off
Tcoupl           =  no
; Pressure coupling is off
Pcoupl           = no
; Initial velocities are read from file
```

6. **Model A: Molecular structure file (cellulose_A.pdb) –** Attached .txt file.

```
------------------------
Filename: cellulose.pdb
------------------------
TITLE     hald
REMARK    THIS IS A SIMULATION BOX
CRYST1   31.931   35.850   41.520  90.00  90.00  86.99 P 1          1
MODEL        1
ATOM      1  C4  GLC     1      2.460   2.740   2.700  1.00  0.00
ATOM      2  C3  GLC     1      3.070   4.080   3.100  1.00  0.00
ATOM      3  O3  GLC     1      4.090   4.510   2.230  1.00  0.00
ATOM      4  HO3 GLC     1      3.750   4.410   1.310  1.00  0.00
ATOM      5  C2  GLC     1      3.680   3.940   4.480  1.00  0.00
ATOM      6  O2  GLC     1      4.260   5.150   4.930  1.00  0.00
ATOM      7  HO2 GLC     1      4.430   5.050   5.890  1.00  0.00
ATOM      8  C6  GLC     1      0.990   0.840   3.480  1.00  0.00
ATOM      9  O6  GLC     1     -0.020   0.790   2.480  1.00  0.00
ATOM     10  HO6 GLC     1     -0.540  -0.030   2.650  1.00  0.00
ATOM     11  C5  GLC     1      1.430   2.310   3.750  1.00  0.00
ATOM     12  O5  GLC     1      1.990   2.270   5.050  1.00  0.00
ATOM     13  C1  GLC     1      2.610   3.460   5.470  1.00  0.00
ATOM     14  O1  GLC     1      3.320   3.140   6.660  1.00  0.00
ATOM     15  C4  GLC     2      2.580   3.260   7.890  1.00  0.00
ATOM     16  C3  GLC     2      1.970   1.910   8.290  1.00  0.00
ATOM     17  O3  GLC     2      0.950   1.490   7.420  1.00  0.00
ATOM     18  HO3 GLC     2      1.300   1.590   6.500  1.00  0.00
ATOM     19  C2  GLC     2      1.370   2.060   9.670  1.00  0.00
ATOM     20  O2  GLC     2      0.780   0.850  10.120  1.00  0.00
ATOM     21  HO2 GLC     2      0.620   0.950  11.080  1.00  0.00
ATOM     22  C6  GLC     2      4.060   5.160   8.670  1.00  0.00
ATOM     23  O6  GLC     2      5.060   5.220   7.680  1.00  0.00
ATOM     24  HO6 GLC     2      5.580   6.040   7.840  1.00  0.00
ATOM     25  C5  GLC     2      3.610   3.700   8.940  1.00  0.00
ATOM     26  O5  GLC     2      3.050   3.730  10.240  1.00  0.00
ATOM     27  C1  GLC     2      2.430   2.530  10.660  1.00  0.00
ATOM     28  O1  GLC     2      1.720   2.860  11.850  1.00  0.00
```

… continues in attached file.

7. Model A: Topology file – cellulose_A.itp – Attached .itp file

```
------------------------------------------
Filename: cellulose.itp
------------------------------------------
;      Make by CHEN Pan
;           On host: panchen.cermav.cnrs.fr
;           At date: Mon Feb 06 16:42:33 2012
;
;           This is your topology file
;           GRowing Old MAkes el Chrono Sweat
;
[ moleculetype ]
; Name            nrexcl
cellulose_chain_periodic      1
[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge       mass  typeB    chargeB     massB
     1       CH1R    1   GLCB    C4     1      0.232     13.019   ; qtot 0.232
     2       CH1R    1   GLCB    C3     2      0.232     13.019   ; qtot 0.464
     3        OA     1   GLCB    O3     2     -0.642    15.9994   ; qtot -0.178
     4        H      1   GLCB    HO3    2      0.41      1.008   ; qtot 0.232
     5       CH1R    1   GLCB    C2     3      0.232     13.019   ; qtot 0.464
     6        OA     1   GLCB    O2     3     -0.642    15.9994   ; qtot -0.178
     7        H      1   GLCB    HO2    3      0.41      1.008   ; qtot 0.232
     8       CH2     1   GLCB    C6     4      0.232     14.027   ; qtot 0.464
     9        OA     1   GLCB    O6     4     -0.642    15.9994   ; qtot -0.178
    10        H      1   GLCB    HO6    4      0.41      1.008   ; qtot 0.232
    11       CH1R    1   GLCB    C5     5      0.232     13.019   ; qtot 0.608
    12        OR     1   GLCB    O5     5     -0.464    15.9994   ; qtot 0.128
    13       CH1R    1   GLCB    C1     5      0.464     13.019   ; qtot 0.36
    14        OE     1   GLCB    O1     5     -0.464    15.9994   ; qtot 0

    15       CH1R    2   GLCB    C4     5      0.232     13.019   ; qtot 0.232
    16       CH1R    2   GLCB    C3     6      0.232     13.019   ; qtot 0.464
    17        OA     2   GLCB    O3     6     -0.642    15.9994   ; qtot -0.178
    18        H      2   GLCB    HO3    6      0.41      1.008   ; qtot 0.232
    19       CH1R    2   GLCB    C2     7      0.232     13.019   ; qtot 0.464
    20        OA     2   GLCB    O2     7     -0.642    15.9994   ; qtot -0.178
    21        H      2   GLCB    HO2    7      0.41      1.008   ; qtot 0.232
    22       CH2     2   GLCB    C6     8      0.232     14.027   ; qtot 0.464
    23        OA     2   GLCB    O6     8     -0.642    15.9994   ; qtot -0.178
    24        H      2   GLCB    HO6    8      0.41      1.008   ; qtot 0.232
    25       CH1R    2   GLCB    C5     9      0.232     13.019   ; qtot 0.608
    26        OR     2   GLCB    O5     9     -0.464    15.9994   ; qtot 0.128
    27       CH1R    2   GLCB    C1     9      0.464     13.019   ; qtot 0.36
    28        OE     2   GLCB    O1     9     -0.464    15.9994   ; qtot 0
…Continues in attached file.
```

8. **Model B: Molecular structure file (cellulose_B.pdb) –** Attached .pdb file.

```
REMARK Box vectors:
REMARK a:   31.13600   0.00000   0.00000
REMARK b:   -3.74196  32.58988   0.00000
REMARK c:    0.00000   0.00000  31.14000
ATOM      1  C4  GLC 1   1     -0.084   0.248  -2.336  1.00  0.00         C
ATOM      2  C3  GLC 1   1      0.442  -1.115  -1.918  1.00  0.00         C
ATOM      3  O3  GLC 1   1      0.035  -2.124  -2.864  1.00  0.00         O
ATOM      4  HO3 GLC 1   1      0.594  -1.955  -3.651  1.00  0.00         H
ATOM      5  C2  GLC 1   1     -0.030  -1.502  -0.536  1.00  0.00         C
ATOM      6  O2  GLC 1   1      0.768  -2.562  -0.003  1.00  0.00         O
ATOM      7  HO2 GLC 1   1      0.584  -2.602   0.955  1.00  0.00         H
ATOM      8  C6  GLC 1   1     -0.662   2.597  -1.587  1.00  0.00         C
ATOM      9  O6  GLC 1   1     -0.003   3.283  -2.637  1.00  0.00         O
ATOM     10  HO6 GLC 1   1     -0.573   4.012  -2.958  1.00  0.00         H
ATOM     11  C5  GLC 1   1      0.052   1.299  -1.279  1.00  0.00         C
ATOM     12  O5  GLC 1   1     -0.525   0.807  -0.055  1.00  0.00         O
ATOM     13  C1  GLC 1   1      0.146  -0.338   0.449  1.00  0.00         C
ATOM     14  O1  GLC 1   1      0.531   0.705  -3.554  1.00  0.00         O
ATOM     15  C4  GLC 2   2      3.262   4.288   0.389  1.00  0.00         C
ATOM     16  C3  GLC 2   2      3.902   2.954   0.794  1.00  0.00         C
ATOM     17  O3  GLC 2   2      3.569   1.895  -0.084  1.00  0.00         O
ATOM     18  HO3 GLC 2   2      3.717   2.227  -0.998  1.00  0.00         H
ATOM     19  C2  GLC 2   2      3.396   2.589   2.174  1.00  0.00         C
ATOM     20  O2  GLC 2   2      3.952   1.356   2.616  1.00  0.00         O
ATOM     21  HO2 GLC 2   2      3.770   1.298   3.579  1.00  0.00         H
ATOM     22  C6  GLC 2   2      2.758   6.641   1.168  1.00  0.00         C
ATOM     23  O6  GLC 2   2      3.364   7.444   0.175  1.00  0.00         O
ATOM     24  HO6 GLC 2   2      3.074   8.372   0.338  1.00  0.00         H
ATOM     25  C5  GLC 2   2      3.595   5.356   1.441  1.00  0.00         C
ATOM     26  O5  GLC 2   2      3.207   4.946   2.739  1.00  0.00         O
ATOM     27  C1  GLC 2   2      3.723   3.705   3.159  1.00  0.00         C
ATOM     28  O1  GLC 2   2      3.830   4.774  -0.837  1.00  0.00         O
…Continues in attached file.
```

9. Model B: Topology file – cellulose_B.top – Attached .top file

```
#include "./gromos53a6carbo.ff/forcefield.itp"

[ moleculetype ]
Other    3

[ atoms ]
   1     CH1R    1   GLC    C4    1   0.23200    13.019
   2     CH1R    1   GLC    C3    2   0.23200    13.019
   3      OA     1   GLC    O3    2  -0.64200    15.9994
   4      H      1   GLC   HO3    2   0.41000     1.008
   5     CH1R    1   GLC    C2    3   0.23200    13.019
   6      OA     1   GLC    O2    3  -0.64200    15.9994
   7      H      1   GLC   HO2    3   0.41000     1.008
   8     CH2     1   GLC    C6    4   0.23200    14.027
   9      OA     1   GLC    O6    4  -0.64200    15.9994
  10      H      1   GLC   HO6    4   0.41000     1.008
  11     CH1R    1   GLC    C5    5   0.23200    13.019
  12      OR     1   GLC    O5    5  -0.46400    15.9994
  13     CH1R    1   GLC    C1    5   0.46400    13.019
  14      OE     1   GLC    O1    5  -0.46400    15.9994
  15     CH1R    2   GLC    C4    6   0.23200    13.019
  16     CH1R    2   GLC    C3    7   0.23200    13.019
  17      OA     2   GLC    O3    7  -0.64200    15.9994
  18      H      2   GLC   HO3    7   0.41000     1.008
  19     CH1R    2   GLC    C2    8   0.23200    13.019
  20      OA     2   GLC    O2    8  -0.64200    15.9994
  21      H      2   GLC   HO2    8   0.41000     1.008
  22     CH2     2   GLC    C6    9   0.23200    14.027
  23      OA     2   GLC    O6    9  -0.64200    15.9994
  24      H      2   GLC   HO6    9   0.41000     1.008
  25     CH1R    2   GLC    C5   10   0.23200    13.019
  26      OR     2   GLC    O5   10  -0.46400    15.9994
  27     CH1R    2   GLC    C1   10   0.46400    13.019
  28      OE     2   GLC    O1   10  -0.46400    15.9994
… Continues in attached file.
```

10. Model B: Python-based codes for generating topology file. –

Attached .py files

- **CONVERT.py**

```
#!/usr/bin/python

import io
import readline
from numpy import *

def search_array(array,value):
# returns array index for value
   for index,item in enumerate(array):
     if item.strip() == value:
       return index

def atnr2res(nr,atoms):
# search array atoms for atom number nr, returns atom residue name

   atom_res=atoms[int(nr)-1].split()[3]
   return atom_res

def atnr2resNr(nr,atoms):
# search array atoms for atom number nr, returns atom residue number as integer

   atom_resnr=atoms[int(nr)-1].split()[2]
   atom_resnr=int(atom_resnr)
   return atom_resnr

def inter(pair1,pair2,atoms,array):

   vW6 = {
   "H":0.,
   "CH0R":6.873E-2,
   "CH1R":6.873E-2,
   "CH2R":6.873E-2,
   "CH0":4.838E-2,
   "CH1":5.396E-2,
   "CH2":6.873E-2,
   "CH3":8.278E-2,
   "HO":0.,
   "OR":4.756E-2,
   "OA":4.756E-2,
   "OE":4.756E-2
```

```
}
vW12 = {
"H":0.,
"CH0R":1.36E-3,
"CH1R":1.36E-3,
"CH2R":1.36E-3,
"CH0":1.837E-3,
"CH1":1.933E-3,
"CH2":2.1775964E-3,
"CH3":2.455782E-3,
"HO":0.,
"OR":0.685E-3,
"OA":1.125E-3,
"OE":1.125E-3
}

nvW6 = {
"HO1O5":0.,
"HO1C3":0.,
"HO1C5":0.,
"HO2C4":0.,
"HO2O5":0.,
"HO3C1":0.,
"HO3C5":0.,
"HO4C2":0.,
"HO4O5":0.,
"C6C1":5.689469E-3,
"C6C3":5.689469E-3,
"O1C3":3.268799E-3,
"O1C5":3.268799E-3,
"O2C4":3.268799E-3,
"O3C1":3.268799E-3,
"O3C5":3.268799E-3,
"O4C2":3.268799E-3,
"O1C6":4.110135E-3,
"O3C6":4.110135E-3,
"O1C2":3.268799E-3
}
nvW12 = {
"HO1O5":0.7E-6,
"HO1C3":0.35E-6,
"HO1C5":0.35E-6,
"HO2C4":0.35E-6,
"HO2O5":0.35E-6,
"HO3C1":0.35E-6,
"HO3C5":0.35E-6,
"HO4C2":0.35E-6,
```

```
   "HO4O5":0.35E-6,
   "C6C1":3.33986E-6,
   "C6C3":3.33986E-6,
   "O1C3":2.5E-6,
   "O1C5":2.5E-6,
   "O2C4":2.5E-6,
   "O3C1":2.5E-6,
   "O3C5":2.5E-6,
   "O4C2":2.5E-6,
   "O1C6":8.4108E-6,
   "O3C6":8.4108E-6,
   "O1C2":2.5E-6
   }
   atom1_name = atoms[int(pair1)-1].split()[4]
   atom2_name = atoms[int(pair2)-1].split()[4]
   atom1_type = atoms[int(pair1)-1].split()[1]
   atom2_type = atoms[int(pair2)-1].split()[1]


# 1. O1C2 between rings
   if ( (atom1_name+atom2_name =='O1C2' and atnr2resNr(pair1,atoms)+1
==atnr2resNr(pair2,atoms)) ):
      array.append('%+5s %+5s   1   %4e %4e; %4s %4s N' %
(str(pair1),str(pair2),nvW6[atom1_name+atom2_name],(nvW12[atom1_name+atom2
_name]), str(atom1_name), str(atom2_name) ))
   else:
# 2. N interactions in ring - a1a2
    if ( nvW6.has_key(atom1_name+atom2_name) and
atnr2resNr(pair1,atoms)==atnr2resNr(pair2,atoms) ):
      array.append('%+5s %+5s   1   %4e %4e; %4s %4s N' %
(str(pair1),str(pair2),nvW6[atom1_name+atom2_name],
(nvW12[atom1_name+atom2_name]), str(atom1_name), str(atom2_name) ))
    else:
# 3. N interactions in ring - a2a1
     if (nvW6.has_key(atom2_name+atom1_name) and
atnr2resNr(pair1,atoms)==atnr2resNr(pair2,atoms) ):
      array.append('%+5s %+5s   1   %4e %4e; %4s %4s N' %
(str(pair1),str(pair2),nvW6[atom2_name+atom1_name],
(nvW12[atom2_name+atom1_name]), str(atom2_name), str(atom1_name) ))
     else:
# 4. standard interactions in ring
       array.append('%+5s %+5s   1   %4e %4e; %4s %4s' %
(str(pair1),str(pair2),vW6[atom1_type]*vW6[atom2_type],vW12[atom1_type]*vW12[
atom2_type], str(atom1_name), str(atom2_name)))

   return array
```

```
def atnr2name(nr,atoms):
#returns atom name and atom type for nr in atoms array
   atom_name=atoms[int(nr)-1].split()[4]
   atom_type=atoms[int(nr)-1].split()[1]
   return atom_name, atom_type


def array_remove(array,exclude):
   newlist = [item for item in array if not item.strip().startswith(exclude)]
   return newlist


lines = []
f = open('topol.top','r')

for line in f:
   lines.append(line)
f.close

atoms_index = search_array(lines,'[ atoms ]')
bonds_index = search_array(lines,'[ bonds ]')
pairs_index = search_array(lines,'[ pairs ]')
exclus_index = search_array(lines,'[ exclusions ]')
angles_index = search_array(lines,'[ angles ]')


head = lines[:(pairs_index+1)]
atoms0 = lines[(atoms_index+1):(bonds_index-1)]
pairs0 = lines[(pairs_index+1):(exclus_index-1)]
exclus0 = lines[(exclus_index+1):(angles_index-1)]
tail = lines[exclus_index:]

atoms = array_remove(atoms0,";")
pairs = array_remove(pairs0,";")
exclus = array_remove(exclus0,";")

pairs_new = []
for index,it in enumerate(pairs):
# LJ interactions for [ pairs ] section
   item = it.strip()
   pair1 = item.split()[0]
   pair2 = item.split()[1]
   inter(pair1,pair2,atoms,pairs_new)

for index,it in enumerate(exclus):
# LJ interactions for [ exclusions ] section
   item = it.strip()
```

```
    if (len(item.split()) == 2):
      pair1 = item.split()[0]
      pair2 = item.split()[1]
      inter(pair1,pair2,atoms,pairs_new)
    if (len(item.split()) == 3):
     pair1=item.split()[0]
     pair2=item.split()[1]
     pair3=item.split()[2]
     inter(pair1,pair2,atoms,pairs_new)
     inter(pair1,pair3,atoms,pairs_new)

#for index,it in enumerate(atoms):
#    item=it.strip()
#    print item

top_new = open('top_new.top','w')

for i,item in enumerate(head):
    top_new.write((item))

for item in pairs_new:
    top_new.write((item+"\n"))

top_new.write("\n")


for item in tail:
    top_new.write((item))
```

- **Process_cif.py**

```
#!/usr/bin/python

from __future__ import print_function
import sys
import math
import copy
import argparse

#-----------------------------------------------#
# Data
#-----------------------------------------------#
atom_ordering =
{'GLC':{'C4':0,'C3':1,'O3':2,'HO3':3,'C2':4,'O2':5,'HO2':6,'C6':7,'O6':8,'HO6':9,'C5':10,'O5'
:11,'C1':12,'O1':13}}
```

```
#-----------------------------------------------#
# Subroutines
#-----------------------------------------------#
def get_coords(cif_data,identifier):

  for i,line in enumerate(cif_data):
    if '_atom_site_disorder_group' in line:
      istart = i+1
    if '_geom_special_details' in line:
      istop = i-1

  raw_coords = cif_data[istart:istop]

  coords = []
  for line in raw_coords:
    dat = line.split()
    symbol = dat[0][0:2]
    chain = int(dat[0][2])
    try:
      conformer = dat[0][3]
    except:
      conformer = identifier
    if conformer != identifier:
      continue
    u = float(dat[2].split('(')[0])
    v = float(dat[3].split('(')[0])
    w = float(dat[4].split('(')[0])
    coords.append([symbol,chain,u,v,w])

  return coords, istop

def box_params(cif_data):

  for line in cif_data:
    if "_cell_length_a" in line: a = float(line.split()[1])
    if "_cell_length_b" in line: b = float(line.split()[1])
    if "_cell_length_c" in line: c = float(line.split()[1])
    if "_cell_angle_alpha" in line: alpha_deg = float(line.split()[1])
    if "_cell_angle_beta"  in line: beta_deg  = float(line.split()[1])
    if "_cell_angle_gamma" in line: gamma_deg = float(line.split()[1])

  alpha = alpha_deg*math.pi/180.0
  beta = beta_deg*math.pi/180.0
  gamma = gamma_deg*math.pi/180.0
  ca = math.cos(alpha)
  cb = math.cos(beta)
  cc = math.cos(gamma); sc = math.sin(gamma)
```

```
  fractional_volume = math.sqrt(1-ca*ca-cb*cb-cc*cc+2*ca*cb*cc)
  # volume = a*b*c*fractional_volume

  av = [a,0.0,0.0]
  bv = [b*cc,b*sc,0.0]
  cv = [c*cb,c*(ca-cb*cc)/sc,c*fractional_volume/sc]

  return [av,bv,cv]

def extend(fractional_coords,box_vectors,rep_a,rep_b,rep_c):

  extended_fractional_coords = []
  for ia in range(0,rep_a):
    a = float(ia)
    for ib in range(0,rep_b):
      b = float(ib)
      for ic in range(0,rep_c):
        c = float(ic)
        for [symbol,chain,u,v,w] in fractional_coords:
          extended_fractional_coords.append([symbol,chain,u+a,v+b,w+c])

  [av,bv,cv] = box_vectors
  avx = [rep_a*ave for ave in av]
  bvx = [rep_b*bve for bve in bv]
  cvx = [rep_c*cve for cve in cv]
  expanded_box_vectors = [avx,bvx,cvx]

  return extended_fractional_coords,expanded_box_vectors

def fractional_to_cartesian(fractional_coords,box_vectors):

  [av,bv,cv] = box_vectors
  cartesian_coords = []
  for [symbol,chain,u,v,w] in fractional_coords:
    x = u*av[0] + v*bv[0] + w*cv[0]
    y = u*av[1] + v*bv[1] + w*cv[1]
    z = u*av[2] + v*bv[2] + w*cv[2]
    cartesian_coords.append([symbol,chain,x,y,z])

  return cartesian_coords

def remove_aliphatic_hydrogens(cartesian_coords):

  # Note that this routine assumes hydrogens are listed directly
  # after atoms they are covalently bonded to - IMPORTANT

  united_atom_coords = []
```

```
  last_heavy_atom = 'C'
  for [symbol,chain,x,y,z] in cartesian_coords:
    current_atom = symbol[:-1]
    atom_number = symbol[-1]
    if current_atom != 'H':
      united_atom_coords.append([symbol,chain,x,y,z])
      last_heavy_atom = current_atom
    else:
      if last_heavy_atom != 'C':
        symbol = current_atom + last_heavy_atom + atom_number
        united_atom_coords.append([symbol,chain,x,y,z])

  return united_atom_coords

def gromacs_ordering(coords,residue):

  reordered_coords = copy.copy(coords)
  new_order = atom_ordering[residue]
  n_at_tot = len(coords); n_at_res = len(new_order)
  if n_at_tot % n_at_res != 0:
    print('Error: do not have whole number of repeat residue units')
    sys.exit()
  else:
    n_rep = n_at_tot/n_at_res
    for i in range(0,n_rep):
      for line in coords:
        symbol = line[0]
        reordered_coords[i*n_at_res+new_order[symbol]] = line

  return reordered_coords

def format_output(cartesian_coords,residue_name,box_vectors):

  [[v1x,v1y,v1z],[v2x,v2y,v2z],[v3x,v3y,v3z]] = box_vectors
  # Note: factor of 0.1 here converts length units from Angstrom to nm
  gro_format = []; pdb_format = []
  gro_format.append("Header, t = 0.0\n")
  gro_format.append("%5d" % len(cartesian_coords) + "\n")
  pdb_format.append("REMARK Box vectors:\n")
  pdb_format.append("REMARK a: " + "%10.5f%10.5f%10.5f" % (v1x, v1y, v1z) + "\n")
  pdb_format.append("REMARK b: " + "%10.5f%10.5f%10.5f" % (v2x, v2y, v2z) + "\n")
  pdb_format.append("REMARK c: " + "%10.5f%10.5f%10.5f" % (v3x, v3y, v3z) + "\n")
  for i,[symbol,chain,x,y,z] in enumerate(cartesian_coords):
    gro_format.append("%5d%-5s%5s%5d%8.3f%8.3f%8.3f%8.4f%8.4f%8.4f" %
(chain,residue_name,symbol,i+1,0.1*x,0.1*y,0.1*z,0.0,0.0,0.0) +"\n")
```

```
    pdb_format.append("%-6s%5d%2s%-
4s%3s%1s%1s%4s%4s%8.3f%8.3f%8.3f%6.2f%6.2f%12s" %
('ATOM',i+1,'',symbol,residue_name,'',chain,chain,'',x,y,z,1.0,0.0,symbol[0]) + "\n")
  gro_format.append("%10.5f%10.5f%10.5f%10.5f%10.5f%10.5f%10.5f%10.5f%10.5f"
% (0.1*v1x,0.1*v2y,0.1*v3z,0.1*v1y,0.1*v1z,0.1*v2x,0.1*v2z,0.1*v3x,0.1*v3y) + "\n")
  return gro_format,pdb_format

#------------------------------------------------#
# Main entry point
#------------------------------------------------#
if __name__ == "__main__":

  parser = argparse.ArgumentParser()
  parser.add_argument("cif_file", help="source data file")
  parser.add_argument("--ident", help="conformer identifier, default = 'A'", default =
'A')
  parser.add_argument("--res", help="GROMACS residue code, default = not
specified", default = '')
  parser.add_argument("--rep_a", help="number of repeats along crystallographic a
axis, default = 1", type=int, default = 1)
  parser.add_argument("--rep_b", help="number of repeats along crystallographic b
axis, default = 1", type=int, default = 1)
  parser.add_argument("--rep_c", help="number of repeats along crystallographic c
axis, default = 1", type=int, default = 1)
  args = parser.parse_args()
  basename = args.cif_file.split('.')[0] + '_' + str(args.rep_a) + 'x' + str(args.rep_b) + 'x' +
str(args.rep_c)
  residue = args.res

  # Read in data from .cif file
  with open(args.cif_file,'r') as f:
    cif_data = f.readlines()

  # Construct Cartesian coordinates from fractional coordinates
  fractional_coords,last_line = get_coords(cif_data,args.ident)
  box_vectors = box_params(cif_data[:last_line])
  fractional_coords,expanded_box_vectors =
extend(fractional_coords,box_vectors,args.rep_a,args.rep_b,args.rep_c)
  cartesian_coords = fractional_to_cartesian(fractional_coords,box_vectors)
  if residue is not '':
    cartesian_coords = remove_aliphatic_hydrogens(cartesian_coords)
    reordered_coords = gromacs_ordering(cartesian_coords,residue)

  # Write out .gro and .pdb files
  gro_out,pdb_out =
format_output(reordered_coords,residue,expanded_box_vectors)
  with open(basename+'.pdb','w') as f: f.writelines(pdb_out)
```

```
with open(basename+'.gro','w') as f: f.writelines(gro_out)
```

11. Model B: Python-based code instructions – Attached .txt file

Steps to prepare input file for GROMACS simulation of crystalline cellulose

0. Download cellulose.cif file as supporting information from:

Y. Nishiyama, P. Langan and H. Chanzy
"Crystal structure and hydrogen-bonding system in cellulose-1beta
from synchrotron X-ray and neutron fiber diffraction"
J. Am. Chem. Soc., 2002, 124(31), 9074-9082

Also download gromos53a6carbo ff from:

W. Plazinki, A. Lonardi and P. H. Hunenberger
"Revision of the GROMOS 56A6_CARBO force field: Improving the description
of ring-conformational equilibria in hexopyranose-based carbohydrates"
J. Comp. Chem., 2015, 37(3), 354-365

And copy gromos53a6carbo.ff directory into current working directory
or system-wide GROMACS ff directory. Also copy CONVERT.py into current
working directory and make it executable:

chmod +x CONVERT.py

1. Run custom-written process_cif.py to convert fractional coordinates in
cif file into Cartesian coordinates in both .gro and .pdb formats,
noting that this code is specialized for cellulose and GROMACS

process_cif.py cellulose.cif A GLC

2. Manually edit cellulose_A.gro to remove one of the glucose units and form

cellulose_A_monomer.gro

3. Run pdb2gmx to generate topology file for monomer and reordered .gro file

pdb2gmx -f cellulose_A_monomer.gro -o cellulose_A_monomer_reordered.gro -ff gromos53a6carbo -water none

4. Run CONVERT.py to modify topology file for carbohydrate FF and overwrite old topol.top file

./CONVERT.py
mv top_new.top topol.top

5. Manually edit new topol.top to specify that there are two separate molecules

in unit cell, changing the value on the final line from 1 to 2.

6. Set up GROMACS control files for energy minimization, equilibration and production runs

7. Check that full set of GROMACS input files are present and correct:

cellulose_UxVxW.gro
topol.top
minimize.mdp
equil_nvt_TEMPERATURES.mdp
run_nve.mdp
gromos53a6carbo.ff (directory)